THE ANNALS *of* MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

Vol. XV, No. 3 — September, 1944

THE ANNALS OF MATHEMATICAL STATISTICS

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore 2, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the ANNALS OF MATHEMATICAL STATISTICS, Mt. Royal & Guilford Aves., Baltimore 2, Md., or to the Secretary of the Institute of Mathematical Statistics, P. S. Dwyer, 116 Rackham Hall, University of Michigan, Ann Arbor, Mich.

Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue. The months of issue are March, June, September and December. Because of war-time difficulties of publication, issues may often be from two to four weeks late in appearing. *Subscribers are therefore requested to wait at least 30 days after month of issue before making inquiries concerning non-delivery.*

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is $5.00 per year. Single copies $1.50. Back numbers are available at $5.00 per volume, or $1.50 per single issue.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879

THE ELEMENTARY GAUSSIAN PROCESSES

By J. L. Doob

University of Illinois

1. Introduction

One of the simplest interesting classes of temporally homogeneous stochastic processes is that for which the distributions of the defining chance variables $\{x(t)\}$ are Gaussian. It is supposed that

(A) if $t_1 < \cdots < t_\nu$, the multivariate distribution of $x(t_1), \cdots, x(t_\nu)$ is Gaussian,[1] and that

(B) this distribution is unchanged by translations of the t-axis.

The process is N-dimensional if $x(t)$ is an N-tuple $x_1(t), \cdots, x_N(t)$. The means $E\{x(t)\}$[2] are independent of t, and will always be supposed to vanish in the following discussion.

The correlation matrix function $R(t)$: $(r_{ij}(t))$ is defined by

$$r_{ij}(t) = E\{x_i(s)x_j(s+t)\}. \tag{1.1.1}$$

This expectation is independent of s, because of condition (B). The matrix function $R(t)$ satisfies the equation

$$r_{ij}(t) = r_{ji}(-t), \qquad i, j = 1, \cdots, N. \tag{1.1.2}$$

It follows that when $t = 0$ the matrix is symmetric:

$$r_{ij}(0) = r_{ji}(0), \qquad i, j = 1, \cdots, N, \tag{1.1.3}$$

and it is also well known that $R(0)$ is non-negative definite. Conditions on the functions $r_{ij}(t)$ necessary and sufficient that $R(t)$ be the correlation matrix function of a stochastic process were found for the case $N = 1$ by Khintchine[3] and for all N by Cramér.[4]

Hypothesis (A), that the process is Gaussian seems at first a restriction so strong that Gaussian processes are unimportant. These processes are, however, of fundamental importance, for the following reasons.

(i) If $R(t)$ is the correlation matrix function of *any* temporally homogeneous stochastic process, there is, according to Khintchine and Cramér, a Gaussian process with this same correlation function. This Gaussian process is uniquely determined by the correlation function (assuming that all first order moments vanish, as usual). Because of this intimate connection between the temporally homogeneous Gaussian processes and the most general temporally homogeneous

---

[1] Singular Gaussian distributions will not be excluded. For example the $x(t_j)$ may all vanish identically.

[2] The expectation of a chance variable x will be denoted by $E\{x\}$.

[3] *Matematische Annalen*, Vol. 109 (1934), p. 608.

[4] *Annals of Math.*, Vol. 41 (1940), pp. 215–230.

processes, it is not surprising that very few facts are known about specifically Gaussian processes, that is facts which are true of temporally homogeneous Gaussian processes, but not of temporally homogeneous processes in general.

(ii) It follows from (i) that in any investigation of temporally homogeneous stochastic processes involving only first and second moments—for example least squares prediction by linear extrapolation—it may be assumed that the variables are Gaussian. Under this hypothesis, the investigator may be helped by the suggestive specialized interpretations possible in the Gaussian case of results which hold in the general case. For example if $N = 1$, the least squares best prediction in the Gaussian case for $x(n + 1)$ in terms of a linear combination of the variables $x(1), \cdots, x(n)$ is the conditional expectation of $x(n + 1)$ for given $x(1), \cdots, x(n)$, which is the least squares best prediction of $x(n + 1)$ in terms of $x(1), \cdots, x(n)$ with no restriction on the functions involved. Thus the linearity of the prediction, which must be part of the hypothesis in the general case, is automatically true in the Gaussian case. There is necessarily a linear least squares best prediction of $x(n + 1)$ in terms of the complete past $\cdots, x(n - 1), x(n)$ since the corresponding conditional expectation is certainly defined in the Gaussian case, and is linear in that case.

(iii) In many applications, there is a real justification for hypothesis (A) that the process is Gaussian. This is so in certain physical studies, for example, because the Maxwell distribution of molecular velocities is Gaussian. Examples will be given below.

The processes discussed in the present paper are all temporally homogeneous Gaussian processes. Most of the theorems will be valid for any temporally homogeneous processes for which the second moments of the variables exist,[5] with the following changes: independent chance variables which are linear combinations of the $x(s)$ will become merely uncorrelated chance variables; the convergence with probability 1 of a series of such chance variables will become merely convergence in the mean; the conditional expectation of one such variable y for given values of others, $y_1, y_2, \cdots$ will become merely the linear approximation $\sum_j a_j y_j$ of y in terms of the y_j which minimizes

$$E\{[y - \sum_j a_j y_j]^2\},$$

that is to say the conditional expectation becomes the least squares linear approximation.

The following theorem and its corollary are fundamental in the study of linear prediction involving infinitely many variables. The results are implicit in much of the work on the subject but do not seem to have been stated explicitly before.

THEOREM 1.2. *Let* $\cdots, x_0, x_1, \cdots$ *be a sequence of one-dimensional Gaussian*

[5] The processes need not even be temporally homogeneous. It is necessary only that $E\{x(s)\}$ and $E\{x(s)\, x(s + t)\}$ be independent of s.

chance variables with the property that if $n_1 < \cdots < n_\nu$, *the multivariate distribution of* $x_{n_1}, \cdots, x_{n_\nu}$ *is Gaussian and that*

$$E\{\cdots, x_{m-1}, x_m ; x_n\} = x_m \tag{1.2.1}$$

[6]

whenever $m < n$. *Then* $E\{x_m\} = a$ *is independent of* m, *and*

$$\cdots \leqq E\{(x_m - a)^2\} \leqq E\{(x_{m+1} - a)^2\} \leqq \cdots . \tag{1.2.2}$$

If the $\{x_n\}$ *are defined for all negative integers,*

$$\lim_{m \to -\infty} x_m = x_{-\infty} \tag{1.2.3}$$

exists with probability 1 *and*

$$\lim_{m \to \infty} E\{(x_{-\infty} - x_m)^2\} = 0. \tag{1.2.4}$$

If the $\{x_n\}$ *are defined for all positive integers, and if the dispersions in* (1.2.2) *form a bounded sequence,*

$$\lim_{m \to \infty} x_m = x_\infty \tag{1.2.3'}$$

exists with probability 1, *and*

$$\lim_{m \to \infty} E\{(x_\infty - x_m)^2\} = 0.^7 \tag{1.2.4'}$$

It follows from (1.2.1) that

$$E\{x_n\} = E\{E\{x_m ; x_n\}\} = E\{x_m\}. \tag{1.2.5}$$

Hence $\cdots = E\{x_0\} = E\{x_1\} = \cdots$. It will be no restriction to assume from now on that

$$\cdots = E\{x_0\} = E\{x_1\} = \cdots = 0.$$

It also follows from (1.2.1) that

$$E\{x_m x_n\} = E\{E\{x_m ; x_m x_n\}\} = E\{x_m E\{x_m ; x_n\}\} = E\{x_m^2\}. \tag{1.2.6}$$

[8]

Using this equation,

$$E\{x_n^2\} = E\{[(x_n - x_m) + x_m]^2\} = E\{(x_n - x_m)^2\} + E\{x_m^2\}, \tag{1.2.7}$$

and the dispersions of the x_n thus form a monotone non-decreasing sequence.

---

[6] The conditional expectation of a chance variable y for given values of a chance variable η will be denoted by $E\{\eta ; y\}$.

[7] Much of this theorem remains true if (1.2.1) is true but only the first moments of the x_n are supposed finite, no other hypothesis being made on their distributions. Cf. Doob, *Am. Math. Soc. Trans.*, Vol. 47 (1940), pp. 458–460.

[8] Alternatively, (1.2.1) implies that $x_n - x_m$ is uncorrelated with x_m. Then $E\{x_m x_n\} = E\{x_m^2 + (x_n - x_m)x_m\} = E\{x_m^2\}$.

Finally, using (1.2.6),

$$E\{(x_{m+1} - x_m)(x_{n+1} - x_n)\} = 0. \tag{1.2.8}$$

The series

$$\sum_m (x_{m+1} - x_m) \tag{1.2.9}$$

is therefore a series of mutually independent chance variables. According to a well known theorem of Kolmogoroff, a sequence of mutually independent chance variables converges with probability 1 if the means and dispersions form a convergent series. The present theorem follows at once from Kolmogoroff's theorem.

COROLLARY. *Let x be a one-dimensional Gaussian chance variable and let*

$$\begin{matrix} \cdots\cdots\cdots \\ x_{01}\,,\ x_{02}\,,\ \cdots \\ x_{11}\,,\ x_{12}\,,\ \cdots \\ x_{21}\,,\ x_{22}\,,\ \cdots \\ \cdots\cdots\cdots \end{matrix} \tag{1.2.10}$$

be sequences of one-dimensional Gaussian chance variables with the property that if $\nu \geqq 1$, the multivariate distribution of $x, x_{m1}, \cdots, x_{m\nu}$ is Gaussian, and suppose that each variable x_{mn} is a member of every later sequence. Then

$$\begin{aligned} \lim_{m\to-\infty} E\{x_{m1}\,, x_{m2}\,, \cdots ; x\} &= x_- \\ \lim_{m\to+\infty} E\{x_{m1}\,, x_{m2}\,, \cdots ; x\} &= x_+ \end{aligned} \tag{1.2.11}$$

exist with probability 1, *and in the mean, and*

$$x_+ = E\{x_{mn}\,,\ m = 0, \pm 1, \cdots,\ n = 1, 2, \cdots ; x\}. \tag{1.2.12}$$

It will first be shown that the sequence $\{x_m\}$, where

$$x_m = E\{x_{m1}\,, x_{m2}\,, \cdots ; x\}, \tag{1.2.13}$$

has the property demanded in the theorem. In fact, from the definition of conditional expectation, the difference $x - x_n$ has expectation zero and is independent of the variables $\{x_{mj}\}$ for $m \leqq n$, and therefore of the variables $\cdots, x_{n-1}, x_n$. Hence

$$(x - x_n) - (x - x_{n+1}) = x_{n+1} - x_n$$

has expectation zero and is independent of the variables $\cdots, x_{n-1}, x_n$. Therefore the sequence $\{x_{n+1} - x_n\}$ is a sequence of mutually independent chance variables with vanishing expectations. This implies (1.2.1) if $m < n$ because

$$\begin{aligned} E\{\cdots, x_m\,; x_n\} = E\left\{\cdots, x_m\,; x_m + \sum_m^{n-1} (x_{j+1} - x_j)\right\} &= E\{\cdots, x_m\,; x_m\} \\ + \sum_m^{n-1} E\{\cdots, x_m\,; x_{j+1} - x_j\} &= x_m. \end{aligned} \tag{1.2.14}$$

Let a be the common value of $E\{x\}, E\{x_m\}$. Since $x - x_m$ is independent of x_m,

$$E\{x_m - a)^2\} + E\{(x - x_m)^2\} = E\{(x - a)^2\}. \tag{1.2.15}$$

Hence the sequence of dispersions of the x_m is bounded and according to Theorem 1.2 the limits x_- and x_+ in (1.2.1) exist with probability 1. Since $x - x_n$ has expectation zero and is independent of x_{mj} for $m \leqq n$, $x - x_+$ also has expectation zero and is independent of x_{mj} for all m, that is (1.2.12) is true.

The simplest non-trivial special case of this theorem is the following: *Let $x_1, x_2, \cdots, x$ be one-dimensional Gaussian chance variables with the property that if $\nu \geqq 1$ the multivariate distribution of $x, x_1, \cdots, x_\nu$ is Gaussian. Then*

$$\lim_{n\to\infty} E\{x_1, \cdots, x_n; x\} = E\{x_1, x_2, \cdots; x\}, \tag{1.2.16}$$

with probability 1, *and this limit is also a limit in the mean.*

As stated, the theorem and corollary are true without the hypothesis that the chance variables concerned are Gaussian. (The existence of second moments must be assumed if the limits are to exist as limits in the mean.) They are stated for Gaussian variables because the proof is simple in that case, and because that is sufficient for the purposes of this paper.

In discussing t.h.G. processes whose parameter t is not restricted to be integral, the usual continuity hypothesis will be made. It will be supposed that $R(t)$ is continuous at $t = 0$:

$$\lim_{t\to 0} [R(t) - R(0)] = -\tfrac{1}{2}\lim_{t\to 0} E\{[x(t) - x(0)]^2\} = 0. \tag{1.3.1}$$

It is then easily concluded that $R(t)$ is everywhere continuous.

In the continuous parameter case, it would be useful to have the conditions on $R(t)$ necessary and sufficient for the continuity in t of the chance variables $x(t)$ and for the existence of the derivative. No set of necessary and sufficient conditions for the continuity of $x(t)$ is known, although the fact of continuity will not be difficult to prove in the special cases to be considered in §4. The conditions for the existence of $x'(t)$ are quite simple, and will be given in Theorem 1.4.

The spectral function of a one-dimensional t.h.G. process will play an essential role in some of the theorems to be discussed below. If $R(n)$ is the correlation function of a one-dimensional t.h.G. process, $R(n)$ can be expressed in either of the forms

$$R(n) = \int_0^\pi \cos n\lambda \, dF(\lambda) \qquad n = 0, \pm 1, \cdots, \tag{1.3.2}$$

$$R(n) = \int_{-\pi}^{\pi} e^{in\lambda} \, dG(\lambda) \qquad n = 0, \pm 1, \cdots, \tag{1.3.2'}$$

where, $F(\lambda)$, called the spectral function of the process, and $G(\lambda)$, called the complex spectral function of the process, are real monotone non-decreasing functions satisfying the following conditions:

$$\begin{aligned} &F(0) = 0 && G(-\pi) = 0 \\ &F(\lambda-) = F(\lambda), \quad 0 < \lambda < \pi, && G(\lambda-) = G(\lambda), \quad -\pi < \lambda < \pi \\ & && G(\lambda) - G(0+) = G(0) - G(-\lambda+) \\ &F(\pi) = G(\pi) \end{aligned} \tag{1.3.3}$$

$$F(\lambda) = G(\lambda) - G(-\lambda+) = 2G(\lambda) - G(0) - G(0+), \quad 0 < \lambda < \pi \tag{1.3.4}$$

$$F'(\lambda) = 2G'(\lambda)^{9}$$

The last equation of course holds only at points where the derivatives exist. The forms (1.3.2), (1.3.2′) are derived trivially from each other. The correlation function determines the spectral functions uniquely, if the latter are supposed to satisfy (1.3.3). In fact, at the points of continuity of $F(\lambda)$, $G(\lambda)$:

$$\begin{aligned} F(\lambda) &= \frac{\lambda R(0)}{\pi} + \frac{2}{\pi}\sum_{1}^{\infty} R(n)\,\frac{\sin n\lambda}{n} \\ G(\lambda) &= \frac{(\lambda+\pi)R(0)}{2\pi} + \frac{1}{2\pi i}\lim_{\nu\to\infty}\sum_{\substack{-\nu \\ n\neq 0}}^{\nu} R(n)\,\frac{e^{in\lambda}}{n}. \end{aligned} \tag{1.3.5}$$

Conversely if any $F(\lambda)$ or $G(\lambda)$ satisfying the stated conditions is used to determine an $R(n)$ by means of (1.3.2) or (1.3.2′), $R(n)$ is the correlation function of a t.h.G. process. The representation of $R(n)$ in terms of $G(\lambda)$ is frequently more convenient than that in terms of $F(\lambda)$, because of the simple properties of the exponential function. The following relation, which will be used below, exhibits the elegance attained by the use of $G(\lambda)$:

$$\begin{aligned} E\{[\sum_{m} c'_m x(m)]\cdot[\sum_{n} c''_n x(n)]\} &= \sum_{m,n} c'_m c''_n R(m-n) \\ &= \int_{-\pi}^{\pi} (\sum_{m} c'_m e^{im\lambda})(\overline{\sum_{n} c''_n e^{in\lambda}})\, dG(\lambda). \end{aligned} \tag{1.3.6}$$

The correlation function of a one-dimensional continuous parameter t.h.G. process can be represented in either of the following forms:

$$R(t) = \int_{0}^{\infty} \cos t\lambda\, dF(\lambda) \tag{1.3.7}$$

$$R(t) = \int_{-\infty}^{\infty} e^{it\lambda}\, dG(\lambda) \tag{1.3.7′}$$

---

[9] H. Wold, *A Study in the Analysis of Stationery Time Series*, Uppsala, (1938), p. 66.

where the spectral function $F(\lambda)$ and the complex spectral function $G(\lambda)$ are monotone non-decreasing and satisfy the conditions

$$F(0) = 0 \qquad G(-\infty) = 0$$

$$F(\lambda-) = F(\lambda), \quad 0 < \lambda < \infty \qquad G(\lambda-) = G(\lambda) \tag{1.3.8}$$

$$G(\lambda) - G(0+) = G(0) - G(-\lambda+)$$

$$F(\infty) = G(\infty)$$

$$F(\lambda) = G(\lambda) - G(-\lambda+) = 2G(\lambda) - G(0) - G(0+), \quad 0 < \lambda < \infty, \tag{1.3.9}$$

$$F'(\lambda) = 2G'(\lambda).$$

The last equation of course only holds at points where the derivatives exist. The correlation function $R(t)$ determines the spectral functions uniquely if the latter are supposed to satisfy (1.3.8). In fact, at the points of continuity of $F(\lambda)$, $G(\lambda)$:

$$F(\lambda) = \frac{2}{\pi}\int_0^\infty R(t)\,\frac{\sin t\lambda}{t}\,dt$$

$$G(\lambda) = \lim_{T\to\infty}\frac{1}{2\pi i}\int_{-T}^{T} R(t)\,\frac{e^{it\lambda}-1}{t}\,dt + G(0). \tag{1.3.10}$$

THEOREM 1.4. *Let $\{x(t)\}$ be the variables of a one-dimensional continuous parameter t.h.G. process with correlation function $R(t)$ and spectral function $F(\lambda)$. If*

$$\int_0^\infty \lambda^2\,dF(\lambda) < \infty \tag{1.4.1}$$

then

(i) *$R'(t)$, $R''(t)$ exist and are continuous, and $R'(0) = 0$;*

(ii) *$x(t)$ is an absolutely continuous function of t, with probability 1;*

(iii) *for each t,*

$$\lim_{h\to 0}\frac{x(t+h-x(t)}{h} = x'(t) \tag{1.4.2}$$

exists, with probability 1, and this convergence is also true in the mean:

$$\lim_{h\to 0} E\left\{\left[\frac{x(t+h)-x(t)}{h} - x'(t)\right]^2\right\} = 0; \tag{1.4.3}$$

(iv) *the $x'(t)$ process is a t.h.G. process, with correlation function $-R''(t)$ and spectral function $\int_0^\lambda \lambda^2\,dF(\lambda)$.*

Conversely if

$$\liminf_{h\to 0}\frac{R(0)-R(h)}{h^2} = \tfrac{1}{2}\liminf_{h\to 0} E\left\{\left[\frac{x(h)-x(0)}{h}\right]^2\right\} < \infty, \tag{1.4.4}$$

then (1.4.1) *is true.*

This theorem is due to Slutsky. The proof will be sketched here, for completeness. (The hypothesis that the process is a Gaussian process is immaterial, since only the second moments are involved in the proof.)

Proof of (*i*). If the integral (1.4.1) exists, $R'(t)$, $R''(t)$ can be obtained by differentiating under the integral sign in (1.2.2):

$$R'(t) = -\int_0^\infty \lambda \sin t\lambda \, dF(\lambda)$$

$$R''(t) = -\int_0^\infty \lambda^2 \cos t\lambda \, dF(\lambda). \tag{1.4.5}$$

Then $R'(t)$, $R''(t)$ are continuous functions, and $R'(0) = 0$.

Proof of (*ii*), (*iii*), (*iv*). The quantity

$$E\left\{\left[\frac{x(t+h_1) - x(t)}{h_1} - \frac{x(t+h_2) - x(t)}{h_2}\right]^2\right\} \tag{1.4.6}$$

can be evaluated in terms of the correlation function $R(t)$, and approaches 0 with h_1, h_2, if the second derivative $R''(t)$ exists. There is therefore a chance variable $y(t)$ to which the difference quotient converges in the mean:

$$\lim_{h\to 0} E\left\{\left[\frac{x(t+h) - x(t)}{h} - y(t)\right]^2\right\} = 0. \tag{1.4.7}$$

The $y(t)$ process is a t.h.G. process. Moreover the equation

$$E\{x(s)x(s+t)\} = R(t) \tag{1.4.8}$$

can be differentiated to give

$$E\{x(s)y(s+t)\} = E\{x(s-t)y(s)\} = R'(t) \tag{1.4.9}$$

and this in turn when differentiated becomes

$$E\{y(s-t)y(s)\} = E\{y(s)y(s+t)\} = -R''(t). \tag{1.4.10}$$

Hence the $y(t)$ process has correlation function $-R''(t)$. Finally,

$$E\left\{\left[x(t) - x(0) - \int_0^t y(s)\, ds\right]^2\right\} = E\{[x(t) - x(0)]^2\}$$
$$+ \int_0^t \int_0^t E\{y(s)y(s')\}\, ds\, ds' - 2\int_0^t E\{[x(t) - x(0)]y(s)\}\, ds = 0, \tag{1.4.11}$$

(evaluating the right side of (1.4.11) in terms of $R(t)$, $R'(t)$, $R''(t)$). Thus $x(t)$ is absolutely continuous, with probability 1, and $y(t)$ is the derived function $x'(t)$. Hence $x'(t)$ exists for almost all t, with probability 1.[10] It follows (Fubini's theorem) that the limit in (1.4.2) exists for each t, with probability 1, except possibly for a t-set of Lebesgue measure 0. Since the process is t.h., the

[10] For the exact meaning and measure-theoretic justification for statements of this type, see Doob, *Am. Math. Soc. Trans.*, Vol. 42 (1937), pp. 107–40.

exceptional set must be either empty or the whole t-line. The exceptional set is therefore empty.

Conversely if (1.4.4) is true, (1.4.1) follows at once from (1.3.7).

It will be convenient to use condensed notation below. If x: $(x_1, \cdots, x_N)$, y: $(y_1, \cdots, y_N)$ are N-dimensional vectors and if A: (a_{ij}) is an N-dimensional square matrix, $x \cdot y$ will denote the matrix $(x_i \bar{y}_j)$, Ax the vector with components $\sum_j a_{ij} x_j$ and (x, y) the number $\sum_i x_i \bar{y}_i$. The adjoint matrix (a^*_{ij}): $a^*_{ij} = \bar{a}_{ji}$ will be denoted by A^*. Throughout this paper, the chance variables will be real-valued, but it will be convenient to use complex constant vectors. The identity matrix will be denoted by I. It will be convenient to denote the i, jth term of the matrix A by $(A)_{ij}$. The following equations will be used frequently:

$$Ax \cdot By = A(x \cdot y)B^*, \qquad (Ax, y) = (x, A^* y).$$

If x is a chance variable, it is clear that $E\{x \cdot x\}$ is a symmetric non-negative definite matrix.

The simplest Gaussian processes are those in which the distribution of future states is based not on the complete past, but only on the immediate present. The precise definition of this (Markoff) property is the following.

(C) If $t_1 < \cdots t_{\nu+1}$ the conditional distribution of $x(t_{\nu+1})$ for given values of $x(t_1), \cdots, x(t_\nu)$ depends only on the value assigned to $x(t_\nu)$. The conditional distribution of $x(t_{\nu+1})$ for given values of $x(t_1), \cdots, x(t_\nu)$ will then be simply the conditional distribution of $x(t_{\nu+1})$ for the assigned value of $x(t_\nu)$.

The processes to be discussed in this paper are the processes with properties (A), (B), (C): temporally homogeneous Gaussian Markoff (t.h.G.M.) processes. The properties of t.h.G.M. processes will also be used to derive properties of the most important simple types of one-dimensional t.h.G. processes—those with rational spectral density functions. Some of the results are contained implicitly in the work of previous writers, but the presentation of the results has in all cases been chosen to stress their specific probability significance, and may therefore appeal even to readers familiar with previous work.

The condition (C) on a Gaussian process is equivalent to the condition (C′) that if $t_1 < \cdots < t_{\nu+1}$

$$E\{x(t_1), \cdots, x(t_\nu); x(t_{\nu+1})\} = E\{x(t_\nu); x(t_{\nu+1})\}. \tag{1.5.1}$$

In fact (C) is at least as strong as (C′). Conversely if (C′) is true,

$$\begin{aligned} x(t_{\nu+1}) &= x(t_{\nu+1}) - E\{x(t_\nu); x(t_{\nu+1})\} + E\{x(t_\nu); x(t_{\nu+1})\} \\ &= y + E\{x(t_\nu); x(t_{\nu+1})\}, \end{aligned} \tag{1.5.2}$$

where y is a Gaussian chance variable with mean 0 uncorrelated with and therefore independent of $x(t_1), \cdots, x(t_\nu)$, and the last term of (1.5.2) is simply a multiple of $x(t_\nu)$. Then the conditional distribution of $x(t_{\nu+1})$ for given $x(t_1), \cdots, x(t_\nu)$ is a Gaussian variable, with mean $E\{x(t_\nu); x(t_{\nu+1})\}$ and dispersion that of y. Since this conditional distribution depends only on $x(t_\nu)$, property (C′)

implies property (C). Hence these properties are equivalent. The condition (C′) can be written in the form

$$(1.5.3)\quad E\{x(\tau),\ \tau \leqq s;\ x(s+t)\} = E\{x(s);\ x(s+t)\},\qquad t > 0.$$

In many applications the stochastic processes either have this property already or will have it if the dimensionality of the processes is increased by the adjunction of auxiliary chance variables. In the latter case the process is called a component process of a t.h.G.M. process. Component processes are discussed in detail below. If a process is a t.h.G. process, the right side of (1.5.3) is a linear transformation (depending only on t) of $x(s)$:

$$(1.5.4)\qquad E\{x(s);x(s+t)\} = A(t)x(s),\qquad t > 0.$$

The matrix function $A(t)$ will be called the transition matrix function. It satisfies the equation (obtained by performing the operation $E\{x(s)\cdot\}$ on both sides of (1.5.4))

$$(1.5.5)\qquad R(t) = R(0)A(t)^*,\qquad t > 0,$$

but is otherwise unrestricted since if (1.5.5) is true, the difference $x(t) - A(t)x(s)$ is uncorrelated with and therefore independent of $x(s)$. In many applications the elements of $R(t)$ will vanish identically except in square matrices down the main diagonal. If this is true, $A(t)$ can also be assumed in this form.

If the variables $\{x(t)\}$ determine an N-dimensional t.h.G.M. process, and if B is a non-singular N-dimensional square matrix, the variables $\{Bx(t)\}$ also determine a t.h.G.M. process. Two processes connected in this way will be called equivalent. If two t.h.G. processes are equivalent, and if one is a Markoff process, the other must be also. If there is a change of variable

$$(1.5.6)\qquad y(t) = Bx(t)$$

taking the t.h.G.M. $x(t)$ process with transition matrix $A(t)$ and correlation matrix $R(t)$ into the equivalent $y(t)$ process with transition matrix $A_1(t)$ and correlation matrix $R_1(t)$, then

$$(1.5.7)\qquad A_1(t) = BA(t)B^{-1},\qquad R_1(t) = BR(t)B^*.$$

If $\{x(t)\}$, $\{y(t)\}$, $\{z(t)\}$ determine t.h.G. processes of dimensions α, β and $\alpha + \beta$ respectively, if the process determined by

$$\{x_1(t), \cdots, x_\alpha(t),\qquad y_1(t), \cdots, y_\beta(t)\}$$

is equivalent to the $z(t)$ process, and if every $x(s)$ is independent of every $y(t)$, the $z(t)$ process will be called the direct product of the $x(t)$ and $y(t)$ processes. The extension of the definition to direct products of more than two processes is clear. If the $x(t)$ and $y(t)$ processes are Markoff processes, their direct product is also a Markoff process. Conversely if the $z(t)$ process is a Markoff process, the factor processes must also be Markoff processes. The following facts about matrices will be used below. If A is any N-dimensional matrix, there is a non-

singular N-dimensional matrix B such that $B^{-1}AB$ is in Jordan canonical form: the elements of $B^{-1}AB$ vanish except for those in certain submatrices down the main diagonal. Each of these submatrices has the form

$$\begin{pmatrix} \lambda & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & \lambda & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & 1 & \lambda \end{pmatrix} \tag{1.5.8}$$

or simply (λ) if it is one-dimensional. The λ's are the characteristic values of A, that is the roots of the characteristic equation $\det. \mid A - \lambda I \mid = 0$, and the sum of the dimensions of the submatrices with a given λ is the multiplicity of λ as a root of this equation. The matrix A is said to have simple elementary divisors corresponding to a given root λ of the characteristic equation if the submatrices in (1.5.7) with that λ are all of dimension 1. Thus orthogonal matrices, symmetric matrices, and skew symmetric matrices have only simple elementary divisors, since they can be put in diagonal form, (with λ's of modulus 1, real, pure imaginary, respectively). The transformation B and the λ's may not be real. If A is real, however, there is a real matrix B such that the elements of $B^{-1}AB$ vanish except for square submatrices down the main diagonal, and the characteristic roots of different submatrices are neither equal nor conjugate imaginary.

The powers of a matrix in Jordan canonical form are easily calculated using the fact that

$$\begin{pmatrix} \lambda & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & \lambda & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & 1 & \lambda \end{pmatrix}^n = \begin{pmatrix} \lambda^n & 0 & \cdot & \cdot & \cdot & 0 \\ n\lambda^{n-1} & \lambda^n & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & n\lambda^{n-1} & \lambda^n \end{pmatrix}. \tag{1.5.9}$$

It follows that in the general case the elements $(A^n)_{ij}$ are linear combinations of λ_j^n, $n\lambda_j^n$, etc., where $\lambda_1, \lambda_2, \cdots$ are the characteristic values of A. Hence if $(A^n)_{ij} \to 0$ as $n \to \infty$, the approach must be exponential. The terms of A^n certainly go to 0 if all the characteristic values of A have modulus less than 1.

The matrix e^A is defined by the usual series formula for the exponential function. If A has the form (1.5.8), e^{tA} can be calculated using (1.5.9):

$$e^{tA} = \begin{pmatrix} e^{t\lambda} & 0 & \cdot & \cdot & \cdot & 0 \\ te^{t\lambda} & e^{t\lambda} & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & te^{t\lambda} & e^{t\lambda} \end{pmatrix}. \tag{1.5.10}$$

It follows that in the general case the elements $(e^{tA})_{ij}$ are linear combinations of $e^{t\lambda_j}$, $te^{t\lambda_j}$, etc. where $\lambda_1, \lambda_2, \cdots$ are the characteristic values of A. If $(e^{tA})_{ij} \to 0$ as $t \to \infty$ the approach must be exponential. The terms of e^{tA} certainly go to 0 if all the characteristic values of A have negative real parts.

A t.h.G.M. process will be called deterministic if the least squares prediction of $x(s+t)$ for given $x(s)$, $(t>0)$, that is $E\{x(s);\, x(s+t)\}$ is always correct:

$$x(s+t) = A(t)x(s) \qquad t>0, \tag{1.6.1}$$

with probability 1.

The following classification of deterministic processes will be useful later. It will be shown that any t.h.G.M. process is the direct product of processes of the following four deterministic types, and of a factor process with no deterministic factors.

$M(0)$. Let $\{x(t)\}$ be the variables of a one-dimensional t.h.G.M. process, with $x(t) = 0$ with probability 1. (The chance variable which is 0 with probability 1 is considered as a Gaussian variable with mean 0 and dispersion 0.) The correlation function of the process vanishes identically.

$M(1)$. A one-dimensional t.h.G.M. process which satisfies

$$x(t) = x(0), \qquad E\{x(t)\} = 0, \qquad E\{x(t)^2\} > 0, \tag{1.6.2}$$

will be called a process of type $M(1)$. The correlation function $R(t)$ is positive and independent of t.

$M(-1)$. A one-dimensional t.h.G.M. process with an integral-valued parameter n, satisfying

$$\cdots = x(-1) = -x(0) = x(1) = \cdots \qquad E\{x(n)\} = 0, \qquad E\{x(n)^2\} > 0 \tag{1.6.3}$$

will be called a process of type $M(-1)$. The correlation function is alternately positive and negative: $R(n) = (-1)^n R(0)$.

$M(e^{i\theta})$. A two-dimensional t.h.G.M. process with

$$E\{x_j(0)\} = 0, \qquad E\{x_j(0)^2\} = \sigma^2 > 0, \qquad E\{x_1(0)x_2(0)\} = 0,$$

$$\begin{aligned} x_1(t) &= x_1(0)\cos t\theta - x_2(0)\sin t\theta \\ x_2(t) &= x_1(0)\sin t\theta + x_2(0)\cos t\theta, \end{aligned} \tag{1.6.4}$$

will be called a process of type $M(e^{i\theta})$. The correlation function is given by

$$R(t) = \begin{pmatrix} \sigma^2\cos t\theta & \sigma^2\sin t\theta \\ -\sigma^2\sin t\theta & \sigma^2\cos t\theta \end{pmatrix}. \tag{1.6.5}$$

A process with variables $\{x(t)\}$ will be called degenerate if there are constants $c_1, \cdots, c_N$ not all 0, such that

$$\sum_j c_j x_j(t) = 0 \tag{1.7.1}$$

with probability 1, for all t. Equation (1.7.1) is true if and only if

$$E\{[\sum_j c_j x_j(t)]^2\} = \sum_{j,k} (R(0))_{jk} c_j c_k = 0 \tag{1.7.2}$$

that is if and only if the correlation matrix $R(0)$ is singular. If a non-degenerate process is a direct product of factor processes, the latter are also non-degenerate. The only degenerate one-dimensional process is that of type $M(0)$.

2. The structure of degenerate and deterministic processes

THEOREM 2.1. *Every degenerate t.h.G.M. process is the direct product of processes of type $M(0)$ and (in some cases) of a non-degenerate t.h.G.M. process.*

In proving this theorem, it can be supposed that the original process has been replaced by an equivalent process, if necessary, so that the symmetric non-negative definite matrix $R(0)$ is in diagonal form, with only 0's and 1's down the main diagonal, say 0 to the νth place and 1 thereafter. Then $x_j(t) = 0$, when $j \leqq \nu$ and the process is obviously the direct product of ν processes of type $M(0)$ and an $(N - \nu)$-dimensional non-degenerate process.

THEOREM 2.2. *Let $\{x(t)\}$ be the variables of a deterministic t.h.G.M. process, with correlation function $R(t)$.*

(i) *The process is the direct product of factor processes of types $M(0)$, $M(\pm 1)$, $M(e^{i\theta})$.*

(ii) *If the parameter t of the process is restricted to the integers, there is a non-singular matrix A such that*

$$x(n) = A^n x(0), \tag{2.2.1}$$

$$R(n) = R(0)A^{*n}, \qquad n = 0, \pm 1, \pm 2, \cdots \tag{2.2.2}$$

$$R(0) = AR(0)A^*. \tag{2.2.3}$$

The transition matrix A is the transform $B0B^{-1}$ of an orthogonal matrix 0. If the process is non-degenerate, A is uniquely determined.

(ii′) *If the parameter of the process runs through all real numbers, there is a matrix Q such that*

$$x(t) = e^{tQ}x(0), \tag{2.2.1′}$$

$$R(t) = R(0)e^{tQ^*}, \qquad -\infty < t < \infty, \tag{2.2.2′}$$

$$QR(0) + R(0)Q^* = 0. \tag{2.2.3′}$$

The matrix Q is the transform BKB^{-1} of a skew symmetric matrix K. If the process is non-degenerate, Q is uniquely determined.

(iii) *Conversely if $R(0)$ is any symmetric non-negative definite matrix, and if $A(Q)$ is any matrix satisfying* (2.2.3) ((2.2.3′)), *where A is non-singular, there is a deterministic t.h.G.M. process with correlation function given by* (2.2.2) ((2.2.2′)) *and satisfying* (2.2.1) ((2.2.1′)).

In proving (i) (ii) and (ii′) it will be permissible to go to processes equivalent to the original one, if convenient. Moreover if the given process can be expressed as a direct product, it will be sufficient to prove (i) (ii) and (ii′) for each factor. Since (i) (ii) and (ii′) are certainly true for processes of type $M(0)$ (with A in (ii) the identity, and Q in (ii′) the null matrix), and since according to Theorem 2.1, processes of type $M(0)$ can be factored out of the given process to leave a non-degenerate remaining factor, if any, it will be sufficient to prove (i) (ii) and (ii′) for non-degenerate processes.

Proof (t integral) of (i) and (ii) for non-degenerate processes. If the process determined by $\{x(n)\}$ is deterministic, (1.6.1) is true. Hence

$$x(\nu + 1) = Ax(\nu). \tag{2.2.4}$$

Then (2.2.1) is true for $n \geqq 0$, and will be established for all n as soon as it is shown that A is non-singular. Using (2.2.1),

$$R(n) = E\{x(0)\cdot x(n)\} = E\{x(0)\cdot A^n x(0)\} = R(0)A^{*n}, \qquad n \geqq 0, \tag{2.2.5}$$

and

$$R(0) = E\{x(1)\cdot x(1)\} = E\{Ax(0)\cdot Ax(0)\} = AR(0)A^*. \tag{2.2.6}$$

Under the present hypotheses, $R(0)$ is non-singular. Then A is determined uniquely by (2.2.5) with $n = 1$, and A cannot be singular because of (2.2.6). There is an equivalent process in which $R(0)$ is the identity. Considering this process, (2.2.6) becomes $I = AA^*$, so that A is orthogonal. Finally there is an equivalent process (obtained by an orthogonal change of variables) in which $R(0)$ is still the identity and the matrix A is in the (real) normal form of orthogonal matrices: all the elements of A are 0 except for two-dimensional rotation matrices or 1's or -1's down the main diagonal. It is now obvious that the process is the direct product of processes of types $M(\pm 1)$, $M(e^{i\vartheta})$.

Proof (t continuously varying) of (i) and (ii') for non-degenerate processes. If the t.h.G.M. process determined by $\{x(t)\}$ is deterministic, (1.6.1) is true. Hence

$$R(t) = E\{x(s)\cdot x(s + t)\} = R(0)A(t)^* \tag{2.2.5'}$$

$$R(0) = E\{x(s + t)\cdot x(s + t)\} = A(t)R(0)A(t)^*. \tag{2.2.6'}$$

The matrix $A(t)$ is uniquely determined by (2.2.5') since $R(0)$ is non-singular. It then follows from (1.6.1) that

$$A(s + t) = A(s)A(t). \tag{2.2.7}$$

The continuity hypothesis (1.3.1) becomes

$$\lim_{t\to 0} R(0)A(t)^* = R(0), \tag{2.2.8}$$

which implies that

$$\lim_{t\to 0} A(t) = I. \tag{2.2.9}$$

It is well known that any solution of (2.2.7) and (2.2.9) can be written in the form $A(t) = e^{tQ}$. If now the right side of (2.2.6') is expanded in powers of t and the coefficient of t is set equal to 0, the resulting equation is (2.2.3'). It can be supposed, going to an equivalent process if necessary, that $R(0)$ is the identity. Then $A(t)A(t)^* = I$, $Q + Q^* = 0$. An equivalent process can be chosen so

that $R(0)$ is still the identity, and so that Q is in the real canonical form of skew symmetric matrices: its elements vanish except for possible two rowed matrices

$$\begin{pmatrix} 0 & \theta \\ -\theta & 0 \end{pmatrix}$$

down the main diagonal. It is now clear that the non-degenerate process is a direct product of factors of type $M(e^{i\theta})$ corresponding to two rowed matrices just described, and factors of type $M(1)$.

Proof of (iii). If $R(0)$ and $A(Q)$ satisfy the conditions of Theorem 1.2 (iii), choose $x(0)$ as any Gaussian variable with correlation matrix $R(0)$. Then define $x(n)$ by (2.2.1) ((2.2.1′)). The resulting stochastic process is temporally homogeneous if and only if $E\{x(s)\cdot x(s+t)\}$ depends only on t. The details of the calculation will be carried out for only for t integral.
In the first place

$$E\{x(n)\cdot x(n+\nu)\} = E\{A^n x(0)\cdot A^{n+\nu}x(0)\} = A^n R(0)A^{*n+\nu}. \tag{2.2.10}$$

Now (2.2.3) can be developed further:

$$R(0) = AR(0)A^* = A^2R(0)A^{*2} = \cdots \tag{2.2.11}$$

so that (2.2.10) reduces to

$$E\{x(n)\cdot x(n+\nu)\} = R(0)A^{*\nu}. \tag{2.2.12}$$

The process is thus temporarally homogeneous, and obviously satisfies the other parts of the definition of a deterministic t.h.G.M. process. Theorem 2.2 is now completely proved.

The restriction imposed on $R(0)$, $A(Q)$ by (2.2.3) ((2.2.3′)) is quite loose. Given $R(0)$, there is always an $A(Q)$ satisfying (2.2.3) ((2.2.3′)) for example the identity (null matrix). Given an A which is the transform of an orthogonal matrix (a Q which is the transform of a skew symmetric matrix) there is always a corresponding $R(0)$: In fact $A(Q)$ can be assumed to be orthogonal (skew symmetric) and the $R(0)$ can be taken as the identity.

3. T.H.G.M. processes with an integral valued parameter

In this section, the parameter t will range through all the integers. The condition (1.5.3) that a t.h.G. process be a t.h.G.M. process can be simplified in the integral parameter case. In fact it will be shown that it is sufficient if

$$E\{\cdots, x(n-1), x(n); x(n+1)\} = E\{x(n); x(n+1)\}, \tag{3.1.1}$$

$$n = 0, \pm 1, \cdots$$

with probability 1. If (3.1.1) is true,

$$x(\nu+1) = Ax(\nu) + \eta(\nu) \tag{3.1.2}$$

where A is the transition matrix of the process and $\eta(\nu)$ has mean 0 and is independent of $\cdots, x(\nu - 1), x(\nu)$. It follows that

$$x(n) = A^{n-m}x(m) + A^{n-m-1}\eta(m) + A^{n-m-2}\eta(m+1) + \cdots + \eta(n-1). \tag{3.1.3}$$

The terms involving the $\eta(j)$ are all independent of $\cdots, x(m-1), x(m)$. This equation therefore implies that

$$E\{\cdots, x(m-1), x(m); x(n)\} = A^{n-m}x(m) = E\{x(m); x(n)\}, \tag{3.1.4}$$

and (3.1.4) is precisely the condition that the process has the Markoff property.

The following lemma will be useful.

LEMMA 3.2. *Let $x:(x_1, \cdots, x_N)$ be any Gaussian chance variable, with $E\{x\} = 0$. Then there is a uniquely determined symmetric non-negative definite matrix S, and a Gaussian chance variable y, such that*

$$E\{y \cdot y\} = I \tag{3.2.1}$$

and

$$x = Sy, \qquad S^2 = E\{x \cdot x\}. \tag{3.2.2}$$

If $x = Sy$, and if S is symmetric, then the second equation in (3.2.2) is certainly true. It is easily seen, by examination of the characteristic values and vectors of the matrix $E\{x \cdot x\}$ that this matrix has a unique symmetric non-negative definite square root S. Hence if there is an S satisfying (3.2.2), there can be only one. The chance variables $x_1, \cdots, x_N$ can be written as linear combinations of N uncorrelated Gaussian chance variables $\xi_1, \cdots, \xi_N$ satisfying $E\{\xi \cdot \xi\} = I$:

$$x = A\xi. \tag{3.2.3}$$

If A is written in the polar form $A = SU$, where S is symmetric and non-negative definite and U is orthogonal, (3.2.3) becomes

$$x = SU\xi = Sy \tag{3.2.4}$$

where $y = U\xi$ satisfies (3.2.1).

It will be shown below that every t.h.G.M. process can be represented as the direct product of factors of certain types. The deterministic types have already been catalogued: $M(0)$, $M(\pm 1)$, $M(e^{i\theta})$. The non-deterministic factor type (integral valued parameter) will now be described.

M. Let $\{\eta(n)\}$ be a sequence of mutually independent N-dimensional Gaussian chance variables with 0 means and a common distribution function. Let A be any N-dimensional square matrix. Define $x(n)$ by

$$x(n) = \sum_{m=0}^{\infty} A^m \eta(n-m) \tag{3.3.1}$$

where it is supposed that[11] A is so chosen that the series converges with probability 1. This will be true, for example, if all the characteristic values of A have modulus less than 1, so that the terms of A^m go exponentially to 0.[12] It will be shown below that it is no restriction to assume that A has this character. The variables $\{x(n)\}$ determine a t.h.G.M. process. Since $x(n)$-$Ax(n-1)$ is independent of $\cdots, x(n-2), x(n-1)$, the $x(n)$ process is a Markoff process with transition matrix A:

$$E\{\cdots, x(n-1); x(n)\} = Ax(n-1). \tag{3.3.2}$$

A process defined in this way will be called a process of type M. A non-singular change of variables $y(n) = Bx(n)$ leads to a process of the same type:

$$y(n) = \sum_{m=0}^{\infty} (BAB^{-1})^m B\eta(n-m). \tag{3.3.3}$$

It will sometimes be convenient to write a process of type M in a form slightly different from (3.3.1). Using Lemma 3.2 it is evident that there are Gaussian variables $\{\xi(n)\}$ satisfying

$$E\{\xi(n)\} = 0, \qquad E\{\xi(m)\cdot\xi(n)\} = \delta_{m,n}I, \qquad m, n = 0, \pm 1, \cdots, \tag{3.3.4}$$

and a symmetric non-negative definite matrix S such that $\eta(n) = S\xi(n)$. Then $S^2 = E\{\xi(n)\cdot\xi(n)\}$ and

$$x(n) = \sum_{m=0}^{\infty} A^m S\xi(n-m). \tag{3.3.5}$$

Under the change of variable $y(n) = Bx(n)$, A becomes BAB^{-1} and S^2 becomes BS^2B^*.

The only condition on A required for convergence in (3.3.5) is that $A^mS \to 0$. It will now be shown that A can always be assumed to have only characteristic values of modulus less than 1, in the sense that there is an A with this property, and satisfying the equations

$$A^mS = \tilde{A}^mS, \qquad m = 1, 2, \cdots. \tag{3.3.6}$$

It is no restriction, going to an equivalent process if necessary, to assume that the elements of A vanish except for those in two square submatrices down the main diagonal, where one submatrix A_1 has only characteristic values of modulus less than 1 and the other, A_2, of modulus greater than or equal to 1. If the matrix S is written in terms of the corresponding submatrices:

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} \qquad S = \begin{pmatrix} S_1 & S_3 \\ S_4 & S_2 \end{pmatrix} \tag{3.3.7}$$

[11] Throughout this paper, if A is any matrix, A^0 is defined as the identity matrix I.

[12] We shall use repeatedly Kolmogoroff's theorem that an infinite series of mutually independent chance variables with zero means converges with probability 1 if the series of their dispersion is convergent. (Kolmogoroff only states the theorem in one dimension, but the extension to n dimension is trivial.) If a series of mutually independent Gaussian variables converges, the series of dispersions converges to the dispersion of the sum.

the condition on A implies that $A_2^m S_2 \to 0$. If it is shown that $S_2 = 0$, it will follow that $S_3 = S_4 = 0$, because S is symmetric and non-negative definite. The matrix $\tilde{A}$ will then be defined by

$$\tilde{A} = \begin{pmatrix} A_1 & 0 \\ 0 & 0 \end{pmatrix} \tag{3.3.8}$$

and A will satisfy (3.3.6) and will have only characteristic values of modulus less than 1. The problem has thus been reduced to the case where A_1 is absent: A only has characteristic values of modulus at least 1, and it must be proved that $A^n S \to 0$ implies that $S = 0$. The proof of this is immediate when A is put into its Jordan canonical form.

The symmetric non-negative definite matrix S satisfies the equations

$$R(0) = \sum_0^\infty A^m S^2 A^{*m} = S^2 + AR(0)A^*. \tag{3.3.9}$$

THEOREM 3.4. *A direct product of processes of type M is also of type M. Conversely any factor process of a process of type M is itself of type M.*

The direct part of the theorem is obvious. To prove the converse, suppose an N-dimensional process of type M has an l-dimensional factor, corresponding to the variables $x_1(n), \cdots, x_l(n)$. It can be supposed, that all factors of type $M(0)$ are separated out, so that there are indices j, k: $1 \leqq j \leqq l \leqq k < N$ such that the $\{x_1(n), \cdots, x_j(n)\}$ and $\{x_{k+1}(n), \cdots, x_N(n)\}$ processes are non-degenerate and that the variables $x_{j+1}(n), \cdots, x_k(n)$ vanish identically. Making a change of variables, if necessary, it can be supposed that $R(0)$ has the form

$$R(0): \begin{pmatrix} j & k-j & N-k \\ I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & I \end{pmatrix} \tag{3.4.1}$$

and that $R(n)$ has the blocks of zeros indicated in (3.4.1). Since $R(1) = R(0)A^*$, A must have the form

$$A: \begin{pmatrix} j & k-j & N-k \\ — & — & 0 \\ 0 & — & 0 \\ 0 & — & — \end{pmatrix}. \tag{3.4.2}$$

Then A^n will have this same form. Finally, because of (3.3.9), S^2, and therefore S must have the form

$$S^2: \begin{pmatrix} — & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & — \end{pmatrix}. \tag{3.4.3}$$

Let A_0 be the matrix whose elements are the same as those of A except that the $(j+1)$th to kth columns of A_0 vanish. Since

$$A_0^m S = A^m S, \qquad m = 0, 1, \cdots \tag{3.4.4}$$

it follows that

$$x(n) = \sum_{m=0}^{\infty} A_0^m S\xi(n-m). \tag{3.4.5}$$

It is now obvious that the $\{x_1(n), \cdots, x_l(n)\}$ process is of type M.

THEOREM 3.5. *A non-degenerate process of type M has no deterministic factors.*

Any factor process is non-degenerate and of type M. To prove the theorem, it will therefore be sufficient to prove that the process itself cannot be deterministic. If it were, we should have

$$x(n) - Ax(n-1) = S\xi(n-1) = 0.$$

Then $S = 0$. But then the process is certainly degenerate, contrary to hypothesis.

THEOREM 3.6. (i) *Every t.h.G.M. process (discrete parameter) is the direct product of processes of type* $M(0)$, $M(\pm 1)$, $M(e^{i\theta})$, M.

(ii) *Let A be a transition matrix of a t.h.G.M. process, with variables* $\{x(n)\}$. *There are mutually independent Gaussian variables* $\cdots, \xi(0), \xi(1), \cdots, \xi$ *satisfying*

$$\begin{aligned} E\{\xi(n)\} &= E\{\xi\} = 0 \\ E\{\xi(n)\cdot\xi(n)\} &= E\{\xi\cdot\xi\} = I, \end{aligned} \tag{3.6.1}$$

and symmetric non-negative definite matrices S, T such that

$$x(n) = \sum_{m=0}^{\infty} A^m S\xi(n-m) + A^n T\xi, \qquad n = 0, 1, \cdots, \tag{3.6.2}$$

$$T^2 = AT^2A^*, \tag{3.6.3}$$

$$R(0) = \sum_{m=0}^{\infty} A^m S^2 A^{*m} + T^2 = AR(0)A^* + S^2, \tag{3.6.4}$$

where the series in (3.6.2) *converges with probability* 1. *If A is non-singular,* (3.6.2) *holds for all n. The sum and last term in* (3.6.2) *are linear transformations of* $x(n)$: (3.6.2) *exhibits in part the decomposition into factor processes described in* (*i*). *The correlation function is given by*

$$\begin{aligned} R(n) &= R(0)A^{*n} \\ R(-n) &= A^n R(0). \end{aligned} \qquad n = 0, 1, \cdots. \tag{3.6.5}$$

(iii) *The transition matrix A is uniquely determined if and only if the process is non-degenerate. In any case, there is a transition matrix whose characteristic values are all of modulus less than or equal to* 1, *and whose characteristic values of modulus* 1 *correspond to simple elementary divisors. The transition matrix A furnishes the solution of the prediction problem of the process*:

$$E\{\cdots, x(m-1), x(m); x(m+n)\} = A^n x(m), \qquad n = 0, 1, \cdots. \tag{3.6.6}$$

The matrix S^2, which is uniquely determined, measures the dispersion of $x(n+1)$ about its predicted value in terms of $x(n)$:

$$E\{[x(n+1) - Ax(n)]^2\} = S^2. \tag{3.6.7}$$

(iv) *Conversely if A is a matrix with at least one characteristic value of modulus less than 1 or of modulus 1 and corresponding to a simple elementary divisor, A is the transition matrix of a t.h.G.M. process, with $R(0)$ not the null matrix. If all the characteristic values of A are as just described, A is the transition matrix of a non-degenerate t.h.G.M. process. If $R(0)$, S, A are matrices satisfying* (3.6.4) *with $R(0)$, S, symmetric and non-negative definite, there is a t.h.G.M. process whose variables can be written in the form* (3.6.2) *with the given $R(0)$, S, A.*

This decomposition of a t.h.G.M. process into deterministic factors can be considered as a special case of the general decomposition theorem of Wold, which is applicable to all t.h.G. process.[13] (Wold only considered the one-dimensional case.) The proof in the present special case is simpler, however, and illuminates the general case.

Proof of (i) and (ii). Equations (1.5.3) and (1.5.4), in the present case, lead to

$$E\{\cdots, x(n-1), x(n); x(n+1)\} = E\{x(n); x(n+1)\} = Ax(n). \tag{3.6.8}$$

The first two terms are equal because the process has the Markoff property. The last term is linear in $x(n)$ because the process is Gaussian. The matrix A can be taken independent of n because the process is temporally homogeneous. Thus (3.6.8) involves the three fundamental properties of the $x(n)$ process. From the definition of conditional expectation, it follows that $x(n+1) - Ax(n)$ is independent of the chance variables $\cdots, x(n-1), x(n)$. Hence the variables

$$\cdots, [x(n) - Ax(n-1)], [x(n+1) - Ax(n)], \cdots$$

are mutually independent. According to Lemma 3.2, there are mutually independent chance variables $\{\xi(n)\}$ satisfying

$$x(n) - Ax(n-1) = S\xi(n), \quad E\{\xi(n)\cdot\xi(n)\} = I, \quad E\{\xi(n)\} = 0, \tag{3.6.9}$$

where S is symmetric, non-negative definite, and satisfies (3.6.7). The matrix S^2 thus measures the dispersion of $x(n)$ about its predicted value $Ax(n-1)$.

The representation (3.6.2) can be obtained very simply. In fact

$$\begin{aligned} x(n) &= [x(n) - Ax(n-1)] + A[x(n-1) - Ax(n-2)] + \cdots \\ &\qquad + A^{n-\nu-1}[x(\nu+1) - Ax(\nu)] + A^{n-\nu}x(\nu) \\ &= \sum_{j=0}^{n-\nu-1} A^j S\xi(n-j) + A^{n-\nu}x(\nu), \end{aligned} \tag{3.6.10}$$

[13] *A Study in the Analysis of Stationery Time Series*, Uppsala (1938), p. 89. See also Kolmogoroff, *Bull. Acad. Sci. URSS Ser. Math.*, Vol. 5 (1941), pp. 3–14 and *Bolletin Moskovskogo Gosudarstvenogo, Mathematika*, Vol. 2 (1941), pp. 1–40), in whose papers the decomposition theorem is brought out in its full significance.

and it will be shown that when $\nu \to -\infty$ (3.6.10) leads to (3.6.2). Before going to the limit, however, we note that in (3.6.10) the sum is independent of the variables $\cdots, x(\nu-1), x(\nu)$, so that

$$E\{\cdots, x(\nu-1), x(\nu); x(n)\} = A^{n-\nu}x(\nu) \tag{3.6.11}$$

which is another way of writing (3.6.6). Moreover, using (3.6.11),

$$R(n-\nu) = E\{x(\nu)\cdot x(n)\} = R(0)A^{*n-\nu}, \tag{3.6.12}$$

which is another way of writing (3.6.5). (The value of $R(n)$ for $n < 0$ is obtained using the fact that $R(-n) = R(n)^*$.) The last term in (3.6.10) is the conditional expectation of $x(n)$ for preassigned $\cdots, x(\nu-1), x(\nu)$. It follows from the corollary to Theorem 1.2 that this conditional expectation converges with probability 1 when $\nu \to -\infty$, but this convergence will be proved directly in the present particular case.

From (3.6.10),

$$E\{x(n)\cdot x(n)\} = R(0) = \sum_{j=0}^{n-\nu-1} A^j S^2 A^{*j} + A^{n-\nu}R(0)A^{*n-\nu}. \tag{3.6.13}$$

The terms of the sum and the last term are all symmetric and non-negative definite matrices. It follows that there is convergence in (3.6.13) when $\nu \to -\infty$:

$$R(0) = \sum_{j=0}^{\infty} A^j S^2 A^{*j} + \lim_{m\to\infty} A^m R(0)A^{*m}. \tag{3.6.14}$$

The convergence of the series of dispersions in (3.6.14) implies that the series of chance variables in (3.6.2) converges, with probability 1. Then when $\nu \to -\infty$ (3.6.10) becomes

$$x(n) = \sum_{j=0}^{\infty} A^j S\xi(n-j) + z(n), \tag{3.6.15}$$

where

$$z(n) = \lim_{\nu\to-\infty} A^{n-\nu}x(\nu). \tag{3.6.16}$$

Since $x(n)$ is independent of $\xi(n+1), \xi(n+2), \cdots$, $z(n)$ is independent of every $\xi(m)$. Moreover, writing $z(0) = T\xi$, where ξ satisfies (3.6.1) and T is symmetric and non-negative definite,

$$z(n) = A^n z(0) = A^n T\xi, \qquad n \geqq 0. \tag{3.6.17}$$

Thus (3.6.3) and (3.6.4) are satisfied. If A is non-singular, (3.6.17) will be correct for negative n also.

The decomposition of the process into factor processes of the types described in the theorem will be obtained by a detailed analysis of the significance of (3.6.2). Under the change of variable $y(n) = Bx(n)$, T^2 becomes BT^2B^*, and

A becomes BAB^{-1}. Making a suitable change of variables, if necessary, it can be supposed that A has the form

$$A: \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} \tag{3.6.18}$$

where the characteristic values of A_1 have moduli unequal to 1 and those of A_2 have modulus 1. The matrix T^2 can be written in terms of submatrices of the same dimensions in a corresponding way:

$$T^2: \begin{pmatrix} T_1^2 & — \\ — & T_2^2 \end{pmatrix} \tag{3.6.19}$$

where T_1, T_2 are symmetric and non-negative definite. A further change of variables may be made, if necessary (transforming only the last n variables) preserving the forms (3.6.18) and (3.6.19) and transforming T_2 into the identity. Then using (3.6.3)

$$A_1 T_1^2 A_1^* = T_1^2, \qquad A_2 A_2^* = I. \tag{3.6.20}$$

Hence A_2 is orthogonal. Developing (3.6.20) further, $A_1^m T_1^2 A_1^{*m} = T_1^2$, for all $m \geqq 0$. When $m \to \infty$ in this equation, the terms in the matrix product on the left involve the mth power of the characteristic values of A_1 (all of modulus different from 1, by hypothesis). Then those characteristic values which actually appear can only be those of modulus less than 1, and the matrix on the left must go to 0 as $m \to \infty$: $T_1 = 0$. Since T is non-negative definite, T must have the simple form

$$T: \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}. \tag{3.6.21}$$

The matrix S can also be divided into corresponding submatrices:

$$S: \begin{pmatrix} S_1 & — \\ — & S_2 \end{pmatrix}. \tag{3.6.22}$$

The convergence of the series in (3.6.4) implies that

$$\lim_{m \to \infty} A_2^m S_2^2 A_2^{*m} = 0.$$

Since A_2 is orthogonal, this means that $S_2 = 0$, and since S is symmetric and non-negative definite, S has the form

$$S: \begin{pmatrix} S_1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{3.6.23}$$

It is now clear from (3.6.2) that the $x(n)$ process is the direct product of a process of type M and a deterministic process corresponding to the division of the variables determining the above submatrices. The deterministic factor process is the direct product of the elementary types already discussed. The variable $z(n)$ and the sum in (3.6.2) are linear transformations of $x(n)$.

Proof of (iii). If the process is non-degenerate, $R(0)$ is non-singular, and the transition matrix is determined uniquely by (3.6.5) with $n = 1$. If the process is degenerate, there will be one or more factor processes of type $M(0)$, and their transition matrices are quite unrestricted. In the non-degenerate case the characteristic values will be of modulus less than 1 (corresponding to a factor of type M, if one is present), or equal to 1 (corresponding to the factors of type $M(\pm 1)$, $M(e^{i\theta})$ making up the deterministic factor, if one is present), and in the latter case the elementary divisors are simple. If the process is degenerate, and if the part of A corresponding to the factors of type $M(0)$ is taken to be the identity, there will be simple elementary divisors corresponding to the characteristic value 1 for each such factor. The remaining statements of (iii) have already been proved.

Proof of (iv). Let A be a matrix with at least one characteristic value of modulus less than 1 or equal to 1 and corresponding to a simple elementary divisor. Then some transform BAB^{-1} has the form

$$\begin{pmatrix} A_1 & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & A_3 \end{pmatrix} \tag{3.6.24}$$

where A_1 (if present) has only characteristic values of modulus less than 1, A_2 (if present) is orthogonal, and both A_1, A_2 are not absent. For the purposes of the present proof it can be supposed that A is already in this form. Define S, T by

$$S: \begin{pmatrix} S_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad T: \begin{pmatrix} 0 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{3.6.25}$$

where the indicated submatrices of S and T are in the same positions as those of A, and where S_1 is any symmetric positive definite (and therefore non-singular) matrix of the proper dimension. The series in (3.6.4) converges and the first equation in (3.6.4) defines a matrix $R(0)$ which obviously satisfies the continued equality. If all the characteristic values of A are as described in the beginning of this paragraph, A_3 can be supposed absent. In this case

$$R(0) = S^2 + T^2 + \cdots$$

is non-singular. The proof of the first two parts of (iv) has now been reduced to that of the last part. Suppose then that $R(0)$, A, S satisfy the hypotheses of the last part of (iv). Then

$$\begin{aligned} R(0) &= AR(0)A^* + S^2 \\ AR(0)A^* &= A^2R(0)A^{*2} + AS^2A^* \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ A^{n-1}R(0)A^{*n-1} &= A^nR(0)A^{*n} + A^{n-1}S^2A^{*n-1}. \end{aligned} \tag{3.6.26}$$

Adding these equations

$$R(0) = \sum_{m=0}^{n-1} A^m S^2 A^{*m} + A^n R(0) A^{*n}. \tag{3.6.27}$$

This equation leads to (3.6.14), and T^2, defined as the limit in (3.6.14), satisfies (3.6.3). Let $\cdots, \xi(-1), \xi(0), \cdots, \xi$ be mutually independent Gaussian variables satisfying (3.6.1). Then the $x(n)$ defined by (3.6.2) determine the variables of a Gaussian process with non-negative values of n, but a slight modification is needed to obtain an expression defined for all n. To obtain this, it can be supposed that A, T, S are in the forms (3.6.24), (3.6.25). Define $\tilde{A}$ by

$$\tilde{A}: \begin{pmatrix} I & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & I \end{pmatrix}. \tag{3.6.28}$$

Then A is orthogonal and $\tilde{A}T = AT$. If now (3.6.2) is used to define $x(n)$ for all n with A^nT replaced by $\tilde{A}^nT$, the $x(n)$ process is a t.h.G.M. process with the desired properties.

The properties of the process reversed in time are of some interest. It is easy to see that if n is replaced by $-n$, a t.h.G.M. process remains a t.h.G.M. process. If the original process is non-degenerate, the new transition matrix is $R(0)A^*R(0)^{-1}$. If the transition matrix remains unchanged when n is replaced by $-n$, $R(0)A^*R(0)^{-1} = A$. This is equivalent to the equation $R(n) = R(-n)$.

The simplest generalization of a t.h.G.M. process is the following. Let the chance variables $\{y(n)\}$ determine a t.h.G. process with the property that for some $N > 0$,

$$E\{\cdots, y(n-1); y(n)\} = E\{y(n-N), \cdots, y(n-1); y(n)\}, \tag{3.7.1}$$

with probability 1. If $N = 1$, the process is a t.h.G.M. process. This type process will be called a t.h.G.M$_N$. process. To avoid notational complications only the one-dimensional case will be considered. The right hand side of (3.7.1) is a linear combination of the variables $y(n-N), \cdots, y(n-1)$. The variables thus satisfy a difference equation of the form

$$y(n) - a_1 y(n-1) - \cdots - a_N y(n-N) = \eta(n) \tag{3.7.2}$$

generalizing (3.6.9), where $\eta(n)$ is independent of the chance variables $\cdots$, $y(n-2)$, $y(n-1)$. The $\{\eta(n)\}$ are mutually independent chance variables with zero means and dispersions independent of n. Equation (3.7.2) leads to

$$y(n) - a_1^{(n-m)} y(m-1) - \cdots - a_N^{(n-m)} y(m-N) = \eta^{(n-m)}(n) \qquad (m \leqq n) \tag{3.7.3}$$

where $\eta^{(n-m)}(n)$ has zero mean and is independent of the chance variables $\cdots$, $y(m-2)$, $y(m-1)$. Hence

$$E\{\cdots, y(m-1); y(n)\} = E\{y(m-N), \cdots, y(m-1); y(n)\}, \qquad m \leqq n. \tag{3.7.4}$$

The difference equation (3.7.2) has been studied in some detail in the past.[14] We shall use an approach which adds insight into the structure of the solution and which clarifies the place of the solution in the general theory of t.h.G. processes. This approach is in terms of N-dimensional t.h.G.M. processes. Define the variables $\{x(n)\}$ of an N-dimensional process by

$$x_j(n) = y(n + j), \qquad n = 0, \pm 1, \cdots, \qquad j = 1, \cdots, N. \tag{3.7.5}$$

The $x(n)$ process is evidently a t.h.G.M. process. If the index N of the $y(n)$ process is the minimum for which (3.7.1) is true, the corresponding $x(n)$ process will be non-degenerate. Then the transition matrix A is uniquely determined, and is evidently

$$A: \begin{pmatrix} 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & 1 \\ a_N & \cdot & \cdot & \cdot & \cdot & \cdot & a_1 \end{pmatrix} \qquad a_1 \neq 0. \tag{3.7.6}$$

The matrix S, measuring the dispersion of the prediction $Ax(n-1)$ of $x(n)$, has the form

$$S: \begin{pmatrix} 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & s \end{pmatrix}. \tag{3.7.7}$$

The characteristic equation of A is simply

$$\alpha^N - a_1\alpha^{N-1} - \cdots - a_N = 0. \tag{3.7.8}$$

The matrix A has only a single characteristic vector corresponding to each characteristic value λ, the vector $(1, \lambda, \cdots, \lambda^{N-1})$. Hence if λ is a multiple root of (3.7.8), it does not correspond to a simple elementary divisor. Therefore, according to Theorem 3.6, all roots of (3.7.8) of modulus 1 must be simple roots. It will be proved below that either no roots have modulus 1 or all roots have modulus 1.

If an N-dimensional non-degenerate t.h.G.M. process is given whose transition matrix A and dispersion matrix S have the forms (3.7.6) and (3.7.7) respectively,

$$x_j(n) - x_{j+1}(n-1) = 0$$

with probability 1, for $j < N$. Then a $y(n)$ process can be defined unambiguously by (3.7.5). Since for fixed j, $x_j(n)$ determines a one-dimensional t.h.G. process, the $y(n)$ process is a t.h.G. process, and (3.7.1) is obviously true, with N minimal if A is non-singular.

Case 1. $S = 0$ (*deterministic case*). In this case the $x(n)$ process is deterministic, and the $y(n)$ process satisfies the equation

$$y(n) = a_1y(n-1) + \cdots + a_Ny(n-N). \tag{3.7.2'}$$

[14] Cf. for example H. Wold, *A Study in the Analysis Of Stationery Time Series*, Uppsala, 1938.

All the roots of (3.7.8) are simple roots, of modulus 1. Since $S = 0$,

$$x(n) = A^n T\xi \qquad n = 0, \pm 1, \cdots \tag{3.7.9}$$

and therefore

$$y(n) = x_1(n-1) = \sum_{j=1}^{N} (A^{n-1}T)_{1j}\xi_j \qquad n = 0, \pm 1 \cdots . \tag{3.7.10}$$

Using either the well known form of the solution of the Nth order difference equation (3.7.2′) or of the powers of an orthogonal matrix, it follows that

$$y(n) = \sum_{j=1}^{N} (\eta_j \cos n\theta_j + \zeta_j \sin n\theta_j) \tag{3.7.11}$$

where the η_j and ζ_j are (one-dimensional) Gaussian variables, and

$$\{\cos \theta_j + i \sin \theta_j\}$$

are the N distinct characteristic values of A, that is the roots of (3.7.8).

Case 2. $S \neq 0$ (*non-deterministic case*). In this case it will now be shown that the $x(n)$ process can have no deterministic factors: that is that the roots of (3.7.8) all have modulus less than 1. In fact let β be a root of (3.7.8), corresponding to the characteristic vector z of A^*:

$$\begin{aligned} z &= (a_N\beta^{N-1}, a_N\beta^{N-2} + a_{N-1}\beta^{N-1}, \cdots, a_N + a_{N-1}\beta + \cdots + a_1\beta^{N-1}) \\ &= (a_N\beta^{N-1}, \cdots, \beta^N). \end{aligned} \tag{3.7.12}$$

Then using (3.6.4),

$$\begin{aligned} (R(0)z, z) &= (AR(0)A^*z, z) + (S^2z, z) \\ &= (R(0)A^*z, A^*z) + (Sz, Sz) \\ &= |\beta|^2(R(0)z, z) + s^2|\beta|^{2N}. \end{aligned} \tag{3.7.13}$$

Hence $|\beta|$ cannot be 1, and the $x(n)$ process can have no deterministic factors. Equation (3.6.2) becomes

$$x(n) = \sum_{m=0}^{\infty} A^n S\xi(n-m) \tag{3.7.14}$$

which leads to

$$y(n) = \sum_{j=1}^{N}\sum_{m=0}^{\infty} (A^mS)_{1j}\xi_j(n-m-1) = s\sum_{m=0}^{\infty} (A^m)_{1N}\xi_N(n-m-1). \tag{3.7.15}$$

According to Theorem 3.6 the only restriction on the coefficients $a_1, \cdots, a_N$ in the two cases $S = 0$, $S \neq 0$, are respectively that equation (3.7.8) has N distinct roots of modulus 1 and all roots of modulus less than 1. Hence (3.7.10) and (3.7.15) furnish (with the stated restrictions on A) the most general t.h.G.M$_N$. processes.

It was shown in Theorem 3.6 that if $R(n)$ is the correlation function of a t.h.G.M. process, $R(n)$ can be expressed in the form (3.6.5), where A is some suitably chosen matrix. Conversely if the correlation function of a t.h.G. process has the form (3.6.5), the process is a t.h.G.M. process since $x(n+1) - Ax(n)$ is then orthogonal to (and therefore independent of) the variables $\cdots, x(n-1), x(n)$. (This fact implies the truth of (3.1.1)). The characterization of t.h.G.M. processes in terms of their correlation functions is thus easily solved. The following theorems characterize one-dimensional t.h.G.M$_N$. processes from various points of view. It will be convenient, and also intrinsically interesting to treat at the same time a slightly larger class of processes: the class of *component processes of t.h.G.M. processes.* A one-dimensional t.h.G. process with variables $\{x_1(n)\}$ will be called a component process of an N-dimensional t.h.G.M. process if there are $N-1$ t.h.G. processes with variables $\{x_2(n)\}, \cdots, \{x_N(n)\}$ such that the N-dimensional process with variables $\{x_1(n), \cdots, x_N(n)\}$ is a t.h.G.M. process. If the variables $\{x(n)\}$ determine an N-dimensional t.h.G.M. process, the t.h.G. processes determined by $\{x_1(n)\}, \cdots, \{x_N(n)\}$ will be called its N component processes. If an $x(n)$ process is not of type $M(0)$ and is a component process of an N-dimensional t.h.G.M. process, it is a component process of a non-degenerate N_1-dimensional t.h.G.M. process, for some $N_1 \leqq N$. It has already been seen that one-dimensional t.h.G.M$_N$. processes are component processes of N-dimensional t.h.G.M. processes.

THEOREM 3.8. *Let $\cdots, x(0), x(1), \cdots$ be the variables of a one-dimensional t.h.G. process. The process is a component process of an N-dimensional t.h.G.M. process if and only if the chance variables*

$$x(0),\ E\{\cdots, x(-1), x(0); x(n)\}, \qquad n = 1, 2, \cdots \tag{3.8.1}$$

are linearly dependent on the first N.

Suppose that the $x(n)$ process is a component process of an N-dimensional $y(n)$ process: $x(n) = y_1(n)$, with correlation function $R_y(n)$ and transition matrix A. Since A satisfies its characteristic equation

$$\det |\, \alpha I - A \,| = \alpha^N - a_1\alpha^{N-1} - \cdots - a_N = 0, \tag{3.8.2}$$

it follows from (3.6.2) that if $\eta(n+N)$ is defined by

$$y(n+N) - a_1 y(n+N-1) - \cdots - a_N y(n) = \eta(n+N) \tag{3.8.3}$$

then $\eta(n+N)$ is independent of $\cdots, y(n-1), y(n)$. Then

$$x(n+N) - a_1 x(n+N-1) - \cdots - a_N x(n) = \eta_1(n+N) \tag{3.8.4}$$

where $\eta_1(n+N)$ is independent of the chance variables $\cdots, x(n-1), x(n)$. Equation (3.8.4) leads to

$$x(n+N+\nu) - a_1^{(\nu)} x(n+N-1) - \cdots - a_N^{(\nu)} x(n) = \eta_1^{(\nu)}(n+N+\nu) \tag{3.8.5}$$

where $\eta_1^{(\nu)}(n+N+\nu)$ has zero mean and is independent of the chance variables $\cdots, x(n-1), x(n)$. If the operator

$$E\{\cdots, x(n-1), x(n); \cdot\}$$

is applied to this equation, the result is

$$E\{\cdots, x(n-1), x(n); x(n+N+\nu)\} - \sum_{m=1}^{N} a_m^{(\nu)} E\{\cdots, x(n-1), x(n); x(n+N-m)\} = 0. \tag{3.8.6}$$

The last term in the sum is $x(n)$ and (3.8.6) is thus the desired linear relation.

Conversely suppose that the $(N+1)$th chance variable in (3.8.1) is linearly dependent on the first N:

$$E\{\cdots, x(n-1), x(n); x(n+N)\} = \sum_{m=1}^{N} a_m E\{\cdots, x(n-1)x(n); x(n+N-m)\}, \quad n = 0, \pm 1, \cdots. \tag{3.8.7}$$

Define the variables $y_1(n), \cdots, y_N(n)$ of an N-dimensional t.h.G. process by

$$\begin{aligned} y_1(n) &= x(n) \\ y_2(n) &= E\{\cdots, x(n-1), x(n); x(n+1) \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_N(n) &= E\{\cdots, x(n-1), x(n); x(n+N-1). \end{aligned} \tag{3.8.8}$$

Then

$$\begin{aligned} &E\{\cdots, y(n-1), y(n); y_1(n+1)\} = y_2(n) \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &E\{\cdots, y(n-1), y(n); y_{N-1}(n+1)\} = y_N(n) \\ &E\{\cdots, y(n-1), y(n); y_N(n+1)\} \\ &\qquad = E\{\cdots, x(n-1), x(n); x(n+N)\} = \sum_{m=1}^{N} a_m y_{N+1-m}(n). \end{aligned} \tag{3.8.9}$$

The $y(n)$ process is therefore a t.h.G.M. process, with transition matrix (3.7.6), and the $x(n)$ process is a component process.

The following particular type of t.h.G. process will be involved in the proof of Theorem 3.9. If the chance variables $\{\eta(n)\}$ determine a t.h.G. process whose correlation function $R_\eta(n)$ vanishes when $n \geqq N$, then according to (1.3.5) the complex spectral function $G_\eta(\lambda)$ of the $\eta(n)$ process is continuous, with derivative $G_\eta'(\lambda)$ given by

$$G_\eta'(\lambda) = \frac{1}{2\pi} \sum_{n=-(N-1)}^{N-1} R_\eta(n). \tag{3.9.1}$$

It is easily verified (using the fact that $R_\eta(n) = R_\eta(-n) = \overline{R_\eta(n)}$) that if α is a root of the equation

$$\sum_{n=-(N-1)}^{N-1} R_\eta(n) z^n = 0, \tag{3.9.2}$$

then $\bar{\alpha}$, $1/\alpha$, $1/\bar{\alpha}$ are also roots, of the same multiplicity. Moreover if $|\alpha| = 1$, α is a root of even multiplicity, since the sum in (3.9.2) is real and non-negative when $|z| = 1$. When $|z| = 1$,

$$|\alpha|\,|(z - \alpha)(z - 1/\bar{\alpha})| = |z - \alpha|^2. \tag{3.9.3}$$

Hence $G'_\eta(\lambda)$ can be written in the following simple form:

$$G'_\eta(\lambda) = |b_0 e^{i(N-1)\lambda} + b_1 e^{i(N-2)\lambda} + \cdots + b_{N-1}|^2 \tag{3.9.4}$$

where the roots of the indicated polynomial have modulus at most 1, and the coefficients are real.

THEOREM 3.9. *Let* $\cdots, x(0), x(1), \cdots$ *be the variables of a one-dimensional t.h.G. process. The process is a component process of a finite-dimensional t.h.G.M. process if and only if the complex spectral function is the sum of the integral of the square of the absolute value of a rational function of* $e^{i\lambda}$ *with real coefficients, and of a monotone non-decreasing function increasing only in a finite number of jumps.*[15] *Specifically*:

(i) *The process is a component process of an N-dimensional t.h.G.M. process if and only if the complex spectral function has the form*

$$G(\lambda) = \int_{-\pi}^{\lambda} \frac{|\beta_0 e^{i(N-1)\lambda} + \cdots + \beta_{N-1}|^2}{|\alpha_0 e^{iN\lambda} + \cdots + \alpha_N|^2}\, d\lambda + \hat{G}(\lambda) \tag{3.9.5}$$

where

(a) $\hat{G}(\lambda)$ *is a monotone non-decreasing function satisfying* (1.3.3), *increasing only by jumps, at no more than N points*;

(b) *the denominator of the integrand vanishes at every discontinuity of* $G(\lambda)$, *and the numerator vanishes at every zero of the denominator, to at least the same order*;

(c) *the coefficients* $\alpha_0, \cdots, \alpha_N, \beta_0, \cdots, \beta_{N-1}$ *are real,* $\alpha_0 \neq 0$, $\beta_0 \neq 0$ *unless the integrand vanishes identically, and the roots of the polynomials in the integrand have modulus less than or equal to* 1.

The integral vanishes identically if and only if the $x(n)$ *process is a component process of an N-dimensional deterministic process, and* $\hat{G}(\lambda)$ *vanishes identically if and only if the variables* $x(n)$ *vanish identically or the* $x(n)$ *process is a component process of an N-dimensional t.h.G.M. process with no deterministic factor.*

(ii) *The process is a t.h.G.M*$_N$. *process (deterministic case) if and only if the*

[15] It is easily seen that the first term of the two can also be described simply as the integral of a rational function of $e^{i\lambda}$, which is non-negative for λ real and is an even function of λ like all complex spectral density functions.

complex spectral function $G(\lambda) = \hat{G}(\lambda)$ *is a monotone non-decreasing function satisfying* (1.3.3) *increasing only in jumps, at no more than* N *points; (non-deterministic case) if and only if the complex spectral function has the form*

$$G(\lambda) = \int_{-\pi}^{\lambda} \frac{d\lambda}{|\, \alpha_0 e^{iN\lambda} + \cdots + \alpha_N \,|^2} \tag{3.9.6}$$

where $\alpha_0, \cdots, \alpha_N$ *are real and* $\alpha_0 \neq 0$.

Proof of (*i*). If the $x(n)$ process is a one-dimensional component of an N-dimensional t.h.G.M. process, it has already been seen that for properly chosen real numbers $a_1, \cdots, a_N$, (3.8.4) is true, where $\eta_1(n + N)$ is independent of the chance variables $\cdots, x(n-1), x(n)$. Equation (3.8.2) can be assumed to have all its roots of modulus less than or equal to 1. It follows from (3.8.4) that $\eta_1(n)$ is independent of $\eta_1(m)$ if $|\, n - m \,| \geqq N$. The complex spectral function of the $\eta_1(n)$ process is therefore continuous, with derivative given by (3.9.4). It will be no restriction to assume that $b_0 \neq 0$ unless the derivative vanishes identically. According to (1.3.6), if $G(\lambda)$ is the complex spectral function of the $x(n)$ process,

$$\begin{aligned} E\{\eta_1(0)\eta_1(n)\} &= \int_{-\pi}^{\pi} e^{i\lambda n} |\, b_0 e^{i(N-1)\lambda} + \cdots + b_{N-1} \,|^2 \, d\lambda \\ &= \int_{-\pi}^{\pi} e^{in\lambda} |\, e^{iN\lambda} - a_1 e^{i(N-1)\lambda} - \cdots - a_N \,|^2 \, dG(\lambda). \end{aligned} \tag{3.9.7}$$

Hence if $\hat{G}(\lambda)$ is the jump function of $G(\lambda)$ ($\hat{G}(-\pi) = 0$, and $\hat{G}(\lambda)$ is constant except for jumps at the same points as those of $G(\lambda)$, and of the same magnitude),

$$\begin{aligned} &\int_{-\pi}^{\lambda} |\, b_0 e^{i(N-1)\lambda} + \cdots + b_{N-1} \,|^2 \, d\lambda \\ &\quad = \int_{-\pi}^{\lambda} |\, e^{iN\lambda} - \cdots - a_N \,|^2 \, d[G(\lambda) - \hat{G}(\lambda)] \\ &\qquad + \int_{-\pi}^{\lambda} |\, e^{iN\lambda} - \cdots - a_N \,|^2 \, d\hat{G}(\lambda). \end{aligned} \tag{3.9.8}$$

Since the first two integrals are continuous in λ, the last must be continuous also. Hence the last integrand must vanish at every discontinuity of $\hat{G}(\lambda)$, that is at every discontinuity of $G(\lambda)$, and the last integral vanishes identically. It follows that

$$G(\lambda) - \hat{G}(\lambda) = \int_{-\pi}^{\lambda} \frac{|\, b_0 e^{i(N-1)\lambda} + \cdots + b_{N-1} \,|^2}{|\, e^{iN\lambda} - \cdots - a_N \,|^2} \, d\lambda \tag{3.9.9}$$

where the numerator vanishes at each zero of the denominator, with the same or greater multiplicity. Since the denominator vanishes at each discontinuity of $G(\lambda)$, there can be at most N discontinuities. If the N-dimensional process is a deterministic process, it can be assumed that all the roots of equation (3.8.2)

have modulus 1, that is that the denominator and hence also the numerator in (3.9.9) have N roots. This can be true only if the numerator vanishes identically: $G(\lambda) \equiv \hat{G}(\lambda)$. If the N-dimensional process has no deterministic factor, it can be assumed that all the roots of equation (3.8.2) have modulus less than 1. Then $G(\lambda)$ can have no discontinuities: $\hat{G}(\lambda) \equiv 0$.

Conversely if $G(\lambda)$ has the form described in Theorem 3.9 (i), $G(\lambda)$ can be assumed in the form (3.9.9) with real coefficients in numerator and denominator and the stated relations between the jumps of $G(\lambda)$ and the zeros of the numerator and denominator in the integrand. (If the integrand vanishes identically and if $G(\lambda)$ has N discontinuities, $a_1, \cdots, a_N$ can be chosen as those numbers making the polynomial

$$e^{iN\lambda} - a_1 e^{i(N-1)\lambda} - \cdots - a_N$$

vanish at the discontinuities of $G(\lambda)$.) Then

$$\begin{aligned} &R(n+N) - a_1 R(n+N-1) - \cdots - a_N R(n) \\ &= \int_{-\pi}^{\pi} e^{i\lambda(n+1)} \frac{[b_0 + \cdots + b_{N-1} e^{i(N-1)\lambda}][b_0 e^{i(N-1)\lambda} + \cdots + b_{N-1}]}{1 - a_1 e^{i\lambda} - \cdots - a_N e^{iN\lambda}} \, d\lambda \\ &\qquad + \int_{-\pi}^{\pi} e^{i\lambda n} [e^{iN\lambda} - \cdots - a_N] \, d\hat{G}(\lambda). \end{aligned} \tag{3.9.10}$$

The last integral vanishes since the bracket vanishes at every jump of $\hat{G}(\lambda)$. The denominator in the first integrand is the value on $|z| = 1$ of a polynomial all of whose roots are outside $|z| = 1$, or on $|z| = 1$. Any zero on $|z| = 1$ corresponds to one of the numerator at the same point. The integral therefore vanishes if $n \geqq 0$ (Cauchy Integral Theorem):

$$R(n+N) - a_1 R(n+N-1) - \cdots - a_N R(n) = 0, \qquad n \geqq 0. \tag{3.9.11}$$

This equation implies that

$$x(n+N) - a_1 x(n+N-1) - \cdots - a_N x(n) \tag{3.9.12}$$

is independent of the chance variables $\cdots, x(n-1), x(n)$, that is that (3.8.4) is true, where $\eta_1(n+N)$ has the stated properties. It has already been seen in the proof of Theorem 3.8 that this implies (3.8.6) and that this in turn implies that the process is a component process of an N-dimensional t.h.G.M. process whose transition matrix A has characteristic equation (3.8.2). In particular if $G(\lambda) \equiv \hat{G}(\lambda)$, the roots of the characteristic equation are of modulus 1, so that the N-dimensional process must be deterministic. If $\hat{G}(\lambda) \equiv 0$, the $x(n)$ process is a component process of an N-dimensional process whose transition matrix A has only characteristic values of modulus less than 1. This N-dimensional process can have no deterministic factors other than one or more of type $M(0)$. If these exist, (and if the $x(n)$ process is not of type $M(0)$) they can be replaced by non-degenerate factors of type M, to obtain an N-dimensional process with no deterministic factor, having the $x(n)$ process as a component process.

Proof of (ii). If the $x(n)$ process is a t.h.G.M$_N$. process, (3.8.4) is true with $\eta_1(m)$ independent of $\eta_1(n)$ if $m \neq n$. The discussion in (i) is therefore simplified by the fact that the numerator in (3.9.9) is constant. If this constant is 0, the spectral function is a function of jumps: $G(\lambda) \equiv \hat{G}(\lambda)$. If this constant is not 0, the denominator in (3.9.9) does not vanish, and $\hat{G}(\lambda)$ therefore vanishes identically. The converse is proved as in (i).

THEOREM 3.10. (i) *If $a_1, \cdots, a_N$ are real numbers, there is a one-dimensional t.h.G. process not of type $M(0)$ with correlation function $R(n)$ satisfying*

$$R(n+N) - a_1R(n+N-1) - \cdots - a_NR(n) = 0 \tag{3.10.1}$$

for $n \geqq 0$ if and only if the equation

$$\alpha^N - a_1\alpha^{N-1} - \cdots - a_N = 0 \tag{3.10.2}$$

has at least one root of modulus less than or equal to 1.

Let $\cdots, x(0), x(1), \cdots$ be the variables of a one-dimensional t.h.G. process not of type $M(0)$.

(ii) *This process is a component process of an N-dimensional t.h.G.M. process if and only if the correlation function $R(n)$ satisfies an Nth order linear difference equation* (3.10.1) *for $n \geqq 0$.*

(iii) *The process is a t.h.G.M$_N$. process if and only if the difference equation* (3.10.1) *is true for $n \geqq -(N-1)$. In this case the vectors $\{x(n), \cdots, x(n+N-1)\}$ determine an N-dimensional t.h.G.M. process.*

(iv) *Equation* (3.10.1) *is satisfied for $n \geqq -N$ if and only if*

$$x(n+N) - a_1x(n+N-1) - \cdots - a_Nx(n) = 0, \quad n = 0, \pm 1, \cdots. \tag{3.10.3}$$

Proof of (ii), (iii), (iv). Let the $x(n)$ process be a component process of an N-dimensional t.h.G.M. $y(n)$ process with correlation function $R_y(n)$: $x(n) = y_1(n)$, and transition matrix A. Since A satisfies its characteristic equation (3.8.2), it follows from (3.6.5) that

$$R_y(n+N) - a_1R_y(n+N-1) - \cdots - a_NR_y(n) = 0, \qquad n \geqq 0. \tag{3.10.4}$$

Then $R(n) = (R_y(n))_{11}$ satisfies this same difference equation. Conversely if (3.10.1) is true for $n \geqq 0$, it has already been proved in the course of the proof of Theorem 3.9 that the $x(n)$ process is a component process of an N-dimensional t.h.G.M. process. This finishes the proof of (ii). Parts (iii) and (iv) are proved similarly.

Proof of (i). According to (ii), if there is a one-dimensional t.h.G. process whose correlation function $R(n)$ satisfies (3.10.1) for $n \geqq 0$, the process is a component process of an N-dimensional t.h.G.M. process whose transition matrix A has (3.10.2) as characteristic equation. Since A has at least one characteristic value of modulus less than or equal to 1, (unless the $x(n)$ process is of type $M(0)$), (3.10.2) must have at least one root of modulus less than or equal to 1. Conversely if (3.10.2) has at least one such root, there is a real N-dimensional matrix A whose characteristic equation is (3.10.2), and which has simple ele-

mentary divisors. According to Theorem 3.6 (ii), A is then the transition matrix of some t.h.G.M. process. The correlation function of this process and hence that of each component process satisfies (3.10.1) for $n \geqq 0$.

THEOREM 3.11. (i) *If $a_1, \cdots, a_N$ are real numbers, there is a one-dimensional t.h.G. process not of type $M(0)$ satisfying*

$$x(n+N) - a_1 x(n+N-1) - \cdots - a_N x(n) = \eta(n+N), \qquad n = 0, \pm 1, \cdots \tag{3.11.1}$$

with $\eta(m)$, $\eta(n)$ independent for $|m-n| \geqq N$ if and only if (3.10.2) has at least one root of modulus less than or equal to 1.

Let $\cdots, x(0), x(1), \cdots$ be the variables of a one-dimensional t.h.G. process.

(ii) *This process is a component process of an N-dimensional t.h.G.M. process if and only if (3.11.1) is true with $\eta(m)$, $\eta(n)$ independent for $|m-n| \geqq N$. In this case $\eta(n+N)$ will be independent of the chance variables $\cdots, x(n-1), x(n)$.*

(iii) *The process is a t.h.G.M. process if and only if in addition to the condition in (ii), $\eta(n)$ is independent of the chance variables $\cdots, x(n-2), x(n-1)$: deterministic case if $\eta(n) = 0$ with probability 1, nondeterministic case otherwise.*

Since this theorem follows readily from the preceding theorems, the proof will be omitted.

The problem of predicting $x(n)$ in terms of $\cdots, x(n-2), x(n-1)$ is trivial (theoretically at least) for t.h.G.M$_N$. processes. In fact these were defined as those processes for which the solution of the prediction problem is simply a linear combination $\sum_{j=1}^{N} a_j x(n-j)$ of the preceding N variables. The solution will now be given for the more general class of component processes of N-dimensional t.h.G.M. processes, processes which have been described from several points of view in the preceding theorems.

The prediction problem for component processes of N-dimensional t.h.G.M. processes will be put into a more general setting. If the one-dimensional chance variables $\{x(n)\}$ determine a t.h.G. process, with correlation function $R(n)$, the problem of finding $E\{\cdots, x(n-2), x(n-1); x(n)\}$ is that of finding a series $\sum_{m=1}^{\infty} \gamma_m x(n-m)$[16] such that

$$x(n) - \sum_{m=1}^{\infty} \gamma_m x(n-m) \tag{3.12.1}$$

is uncorrelated with every $x(n-\nu)$ $(\nu > 0)$:

$$R(\nu) - \sum_{m=1}^{\infty} \gamma_m R(\nu - m) = 0, \qquad \nu > 0. \tag{3.12.2}$$

[16] We are neglecting all convergence difficulties. They become trivial for the applications to be made below.

If the complex spectral function is $G(\lambda)$, (3.12.2) becomes

$$\int_{-\pi}^{\pi} e^{i\nu\lambda}\left\{1 - \sum_{m=1}^{\infty} \gamma_m e^{-im\lambda}\right\} dG(\lambda) = 0, \qquad \nu = 0. \tag{3.12.3}$$

Let $G(\lambda)$ be the integral of its derivative $G'(\lambda)$, that is let $G(\lambda)$ be absolutely continuous. According to (3.12.3) the problem reduces to that of finding a function

$$f(z) = 1 - \frac{\gamma_1}{z} - \frac{\gamma_2}{z^2} - \cdots \qquad z = e^{i\lambda} \tag{3.12.4}$$

such that $f(z)G'$ is of power series type, a function corresponding to an expansion in non-negative powers of z. The dispersion of the error of the prediction is

$$\begin{aligned} E\left\{\left[x(n) - \sum_{m=1}^{\infty} \gamma_m x(n-m)\right]^2\right\} &= \int_{-\pi}^{\pi} \left| e^{in\lambda} - \sum_{m=1}^{\infty} \gamma_m e^{i(n-m)\lambda} \right|^2 dG(\lambda) \\ &= \int_{-\pi}^{\pi} |f|^2 \, dG(\lambda). \end{aligned} \tag{3.12.5}$$

In particular if the $x(n)$ process is a component process of an N-dimensional t.h.G.M. process, $G(\lambda)$ is given by (3.9.5). It will be supposed throughout the following that $\hat{G} \equiv 0$. Then

$$G'(\lambda) = \frac{z(\beta_0 z^{N-1} + \cdots + \beta_{N-1})(\beta_0 + \cdots + \beta_{N-1} z^{N-1})}{(\alpha_1 z^N + \cdots + \alpha_N)(\alpha_0 + \cdots + \alpha_N z^N)}, \qquad z = e^{i\lambda}. \tag{3.12.6}$$

In this case $f \equiv 1$ if $G' \equiv 0$, and otherwise f is given by

$$f(z) = \frac{\beta_0(\alpha_0 z^N + \cdots + \alpha_N)}{\alpha_0 z(\beta_0 z^{N-1} + \cdots + \beta_{N-1})} \tag{3.12.7}$$

so that

$$|f(z)|^2 = \frac{\beta_0^2}{\alpha_0^2 G'}. \tag{3.12.8}$$

The dispersion of the prediction error is $R(0)$ if $G' \equiv 0$ and otherwise is $2\pi\beta_0^2/\alpha_0^2$. The prediction formula for $x(n)$ in terms of the variables $\cdots$, $x(n-\nu-1)$, $x(n-\nu)$ has now been derived for $\nu = 1$, for the chance variables under discussion in this section. The solution for general ν is easily obtained.

As $\nu \to \infty$, the prediction converges with probability 1, according to the corollary to Theorem 1.2. If the process is a component process of an N-dimensional t.h.G.M. process, and if $\hat{G} \equiv 0$ in (3.9.5), the limit is 0. That is, in this case, the best predicted value of $x(n)$ in terms of the distant past is near $E\{x(n)\} = 0$, the same predicted value which would be used with no knowledge of the past.

4. Processes whose parameter t varies continuously

The basic process in terms of which t.h.G.M. processes without deterministic factors were expressed in section 3 was a process whose variables $\{\xi(n)\}$ were Gaussian, with

$$E\{\xi(n)\} = 0, \qquad E\{\xi(m)\cdot\xi(n)\} = \delta_{m,n}I. \tag{4.1.1}$$

The corresponding process in the continuous parameter case is not obtained by replacing the integral parameters m, n in (4.1.1) by continuous parameters. In fact the process so defined does not satisfy any useful continuity conditions. In the present discussion, sums like $\sum_m A_m\xi(m)$ will be replaced by Stieltjes integrals $\int A(t)d\xi(t)$, and $d\xi(t)$ thus will correspond to $\xi(n)$. The $\xi(t)$ process is defined as follows. For any $t_1 < \cdots < t_n$, the chance variables

$$\xi(t_2) - \xi(t_1), \cdots, \xi(t_n) - \xi(t_{n-1})$$

are mutually independent N-dimensional Gaussian chance variables, and if $s < t$,

$$E\{\xi(t) - \xi(s)\} = 0, \qquad E\{[\xi(t) - \xi(s)]\cdot[\xi(t) - \xi(s)]\} = (t - s)I. \tag{4.1.2}$$

This process, called simply a ξ-process below has been discussed in great detail by Bachelier, Wiener and Lévy. The function $\xi(t)$, considered as a function of t is known to be continuous with probability 1.[17] The derivative $\xi'(t)$ does not exist, since $E\{[\xi_i(t + h) - \xi_i(t)]^2\}$ is proportional to h, whereas this mean would be proportional to h^2 if $\xi'(t)$ existed. In fact it has been shown that $\xi(t)$ is (with probability 1) not even of bounded variation in any finite interval. However, if $f(t)$ is a function defined and continuous for $a \leqq t \leqq b$ (where a or b or both may be infinite), the integral

$$\int_a^b f(t)\, d\xi(t) \tag{4.1.3}$$

can be defined as the limit in the mean of the usual Stieltjes sum. If $f(t)$ has a continuous derivative, the integral in (4.1.3) can be evaluated by integration by parts:

$$\int_a^b f(t)\, d\xi(t) = f(b)\xi(b) - f(a)\xi(a) - \int_a^b \xi(t)f'(t)\, dt. \tag{4.1.4}$$

Integrals of the following type will be used below:

$$y(t) = \int_a^t f(t - \tau)\, d\xi(\tau) = f(0)\xi(t) - f(t - a)\xi(a) + \int_a^t \xi(\tau)f'(t - \tau)\, d\tau \tag{4.1.5}$$

[17] Paley and Wiener, "Fourier transforms in the complex domain," *Am. Math. Soc. Colloq. Pub.*, Vol. 19, p. 148.

where $f(t)$ is continuous and has two continuous derivatives. It is then evident that $y(t)$ is continuous, but that $y'(t)$ exists if and only if $f(0) = 0$. If $f(0) = 0$, $y'(t)$ is given by

$$y'(t) = \int_a^t f'(t - \tau)\, d\xi(\tau). \tag{4.1.6}$$

A more general process will also come into the discussion below, and will be called a ζ-process. The chance variables $\{\zeta(t)\}$ of a ζ-process are Gaussian, and have the same independence property as the variables of a ξ-process. The second equation of (4.1.2) is dropped, so that (4.1.2) is replaced by

$$E\{\zeta(t) - \zeta(0)\} = 0, \qquad E\{[\zeta(t) - \zeta(0)]\cdot[\zeta(t) - \zeta(0)]\} = D(t), \tag{4.1.7}$$

where the symmetric and non-negative definite matrix $D(t)$ will sometimes be supposed to have special properties, such as continuity in t, etc. The independence property of the ζ-process implies that

$$E\{[\zeta(t) - \zeta(s)]\cdot[\zeta(t) - \zeta(s)]\} = D(t) - D(s). \tag{4.1.8}$$

Hence $D'(t)$ (if this derivative exists) is symmetric and non-negative definite.

THEOREM 4.1. *If the dispersion matrix $D(t)$ of a ζ-process is continuous, the functions $\{\zeta(t)\}$ are continuous in t, with probability* 1.

The component processes of a ζ-process with a continuous dispersion function are also ζ-processes with continuous dispersion functions. Hence it will be sufficient to prove the theorem in the one-dimensional case. In this case $D(t)$ is non-negative and monotone non-decreasing, according to (4.1.7) and (4.1.8). It can be supposed that $D(t)$ does not vanish identically. Let $D_1(t)$ be an inverse function of $D(t)$: $D[D_1(t)] = t$. Then $\xi(t) = \zeta[D_1(t)]$ defines a ξ-process, and the continuity of $\xi(t)$ implies that of $\zeta(t)$.

The integrals of type (4.1.3) are defined for ζ-processes as for ξ-processes, and satisfy the equations

$$\begin{aligned} E\left\{\int_a^b f(t)\, d\zeta(t)\right\} &= 0 \\ E\left\{\int_a^b f(t)\, d\zeta(t)\cdot\int_a^b g(t)\, d\zeta(t)\right\} &= \int_a^b f(t)g(t)D'(t)\, dt \\ E\left\{\int_a^b A(t)\, d\zeta(t)\cdot\int_a^b B(t)\, d\zeta(t)\right\} &= \int_a^b A(t)D'(t)B(t)^*\, dt, \end{aligned} \tag{4.1.9}$$

where f, g are numerically valued functions and A, B are matrix functions.[18]

The ζ-processes lie at the basis of t.h.G. processes. To every t.h.G. process (discrete parameter) with variables $\{x(n)\}$ correspond two one-dimensional ζ-processes with variables $\{\zeta_1(t)\}$, $\{\zeta_2(t)\}$ such that

$$x(n) = \int_{-\pi}^{\pi} \cos n\lambda\, d\zeta_1(\lambda) + \sin n\lambda\, d\zeta_2(\lambda) \tag{4.1.10}$$

[18] These equations are easily proved using the fact that each integral can be approximated by the usual Riemann-Stieltjes sums.

where the two ζ-processes are mutually independent in the sense that every $\zeta_1(\lambda_1)$ is independent of every $\zeta_2(\lambda_2)$ and where, if $G(\lambda)$ is the complex spectral function of the process,

$$E\{\zeta_i(\lambda)^2\} = G(\lambda). \tag{4.1.11}$$

In the continuous parameter case (4.1.11) becomes

$$x(t) = \int_{-\infty}^{\infty} \cos t\lambda \, d\zeta_1(\lambda) + \sin t\lambda \, d\zeta_2(\lambda). \text{ [19]} \tag{4.1.12}$$

This theorem of Cramér shows that $x(n)$, or $x(t)$ as the case may be, is the limit of a sum of sines and cosines, with Gaussian chance variables as coefficients. The dispersion of each coefficient, which measures the intensity of the corresponding periodic term of the sum, is determined by the spectral function of the process. In particular, if the spectral function $F(\lambda)$ is the integral of its derivative $F'(\lambda)$, each integrand involving $d\zeta_i(\lambda)$ in the above equations can be replaced by one involving $\sqrt{F'(\lambda)}d\xi_i(\lambda)$ where $\xi_i(\lambda)$ is the variable of a ξ-process. Thus in many important cases the processes can be written in a simple way in terms of ξ-processes.

It will be shown below that every t.h.G.M. process can be represented as the direct product of factors of certain types. The deterministic types have already been catalogued: $M(0)$, $M(1)$, $M(e^{i\theta})$. The standard non-deterministic type, as in the discrete parameter case, will be called type M.

M. Let $\{\xi(t)\}$ be the chance variables of an N-dimensional ξ-process, as described above. Let Q be an N-dimensional square matrix, and let S be an N-dimensional symmetric non-negative definite matrix. Define $x(t)$ by

$$x(t) = \int_0^{\infty} e^{sQ} S \, d\xi(t - s) = \int_{-\infty}^{t} e^{(t-\tau)Q} \, d\xi(\tau) \tag{4.2.1}$$

where it is supposed that the improper integrals converge with probability 1. There will be convergence, for example, if Q has only characteristic values with negative real parts so that the elements in the matrix e^{sQ} go to 0 exponentially as $s \to \infty$. (Cf. section 1.) It will be shown below that it is no restriction to assume that Q has this character. The $x(t)$ process is evidently a t.h.G. process. If $u < t$, the chance variable

$$x(t) - e^{(t-u)Q} x(u) = e^{tQ} \int_u^t e^{-sQ} S \, d\xi(s) \tag{4.2.2}$$

is independent of $x(v)$ for $v \leqq u$, since $x(v)$ is expressed in terms of $\xi(s)$ for $s \leqq v$. Therefore the $x(t)$ process is a Markoff process with transition matrix $A(t) = e^{tQ}$:

$$E\{x(v), v \leqq u; x(t)\} = e^{(t-u)Q} x(u), \qquad u < t. \tag{4.2.3}$$

[19] H. Cramér, *Arkiv For Matematik, Astronomi och Fysik*, Vol. 28B, No. 12, pp. 1–17. Cramér only discusses the continuous parameter case, but the other requires no change of method. He allows complex-valued ξ-processes, in terms of which (4.1.10) and (4.1.12) assume a more elegant form.

A process defined in this way will be called a process of type M. A change of variable $y(t) = Bx(t)$ leads to a process of the same type:

$$y(t) = \int_0^\infty e^{sBQB^{-1}} BS\, d\xi(t-s). \tag{4.2.4}$$

The matrix Q goes into BQB^{-1} and if S_1O is the polar form of BS, where S is symmetric and non-negative definite and O is orthogonal, S goes into S_1. (We are using the fact that $O\xi(t)$ defines a second ξ-process.) The correlation function of a process of type M is easily calculated:

$$R(0) = \int_0^\infty e^{sQ} S^2 e^{sQ^*}\, ds, \qquad R(t) = R(0)e^{tQ^*}. \tag{4.2.5}$$

The only condition imposed on Q is that the improper integrals in (4.2.1) converge. This condition is easily seen to be equivalent to the convergence of the integral in (4.2.5). This in turn is equivalent to the condition that

$$\lim_{s\to\infty} e^{sQ} S = 0. \tag{4.2.6}$$

This condition is certainly satisfied if the characteristic values of Q all have negative real parts, and it can always be assumed that this is so. (Cf. the corresponding discussion of processes of type M in the discrete parameter case.)

The analogues in the continuous parameter case of Theorems 3.4 and 3.5 are true. The proofs are substantially the same as the proofs in the discrete parameter case, and will be omitted.

THEOREM 4.3. (i) *Every t.h.G.M. process (continuous parameter) is the direct product of processes of type $M(0)$, $M(1)$, $M(e^{i\theta})$, M.*

(ii) *If $x(t)$ are the variables of such a process, there is a matrix Q such that $A(t) = e^{tQ}$ is a transition matrix function. There is a ξ-process, a Gaussian variable ξ, independent of the $\xi(t)$, satisfying*

$$E\{\xi\} = 0, \qquad E\{\xi\cdot\xi\} = I \tag{4.3.1}$$

and symmetric non-negative definite matrices S, T such that

$$\begin{aligned} x(t) &= \int_0^\infty e^{sQ} S\, d\xi(t-s) + e^{tQ} T\xi \\ &= \int_{-\infty}^t e^{(t-s)Q} S\, d\xi(s) + e^{tQ} T\xi = \int_0^t e^{(t-s)Q} S\, d\xi(s) + e^{tQ} x(0), \end{aligned} \tag{4.3.2}$$

$$QT^2 + T^2Q^* = 0, \tag{4.3.3}$$

$$R(0) = \int_0^\infty e^{sQ} S^2 e^{sQ^*}\, ds + T^2 \tag{4.3.4}$$

$$QR(0) + R(0)Q^* = -S^2, \tag{4.3.5}$$

where the integrals in (4.3.2) converge with probability 1. The integral and the last term in each pair in (4.3.2) are linear transformations of $x(t)$: (4.3.2) exhibits

in part the decomposition into factor processes described in (i). The correlation function is given by

$$R(t) = R(0)e^{tQ^*} \qquad t \geqq 0. \tag{4.3.6}$$
$$R(-t) = e^{tQ}R(0)$$

(iii) *The matrix Q is uniquely determined if and only if the process is non-degenerate. In any case there is a Q whose characteristic values all have negative or zero real parts and whose characteristic values with zero real parts correspond to simple elementary divisors. The matrix Q furnishes the solution to the prediction problem of the process:*

$$E\{x(s),\ s \leqq t;\ x(t+u)\} = e^{uQ}x(t), \qquad u > 0. \tag{4.3.7}$$

The matrix S, which is uniquely determined, measures the dispersion of $x(t)$ about its predicted value:

$$E\{[x(t+u) - e^{uQ}x(t)]^2\} = R(0) - e^{uQ}R(0)e^{uQ^*} \sim uS^2 \qquad (u \to 0). \tag{4.3.8}$$

(iv) *Conversely if Q is a matrix with at least one characteristic value with negative real part or with zero real part and corresponding to a simple elementary divisor, e^{tQ} is the transition matrix function of a t.h.G.M. process with $R(0)$ not the null matrix. If all the characteristic values of Q are as just described, e^{tQ} is the transition matrix function of a non-degenerate t.h.G.M. process. If $R(0)$, S, Q are matrices satisfying (4.3.5) with $R(0)$, S symmetric and non-negative definite, there is a t.h.G.M. process whose variables can be written in the form (4.3.2) with the given $R(0)$, S, Q.*

The proof of Theorem 4.3 follows closely that of Theorem 3.6, and the details will not be given, except as they differ from those of the earlier proof.

Proof of (i). Suppose that the $\{x(t)\}$ are the variables of a t.h.G.M. process which is non-degenerate. The transition matrix function $A(t)$ is then uniquely determined by (1.5.5). Take the conditional expectation of both sides of (1.5.4) for given $x(0)$:

$$A(s+t)x(0) = A(t)A(s)x(0) \qquad s, t > 0. \tag{4.3.9}$$

Since the process is non-degenerate,

$$A(s+t) = A(s)A(t) \qquad s, t > 0. \tag{4.3.10}$$

According to (1.3.1) and (1.5.5)

$$\lim_{t \to 0} R(t) = \lim_{t \to 0} R(0)A(t)^* = R(0), \qquad t > 0. \tag{4.3.11}$$

Hence

$$\lim_{t \to 0} A(t) = I. \tag{4.3.12}$$

It has already been noted that any solution to (4.3.10) under the continuity hypothesis (4.3.12) can be written in the form

$$A(t) = e^{tQ}, \tag{4.3.13}$$

where

$$Q = \lim_{t \to 0} \frac{A(t) - I}{t}.$$

Under a change of variables $y(t) = Bx(t)$, $A(t)$ becomes $BA(t)B^{-1}$ and Q becomes BQB^{-1}. According to Theorem 2.1, if the $x(t)$ process is degenerate, it is the direct product of one or more factors of type $M(0)$ and (perhaps) of a non-degenerate factor. The matrix Q of a factor of type $M(0)$ can be taken as the null matrix. Then the form (4.3.13) is admissible for any t.h.G.M. process, although Q will only be uniquely determined if the process is non-degenerate. Define $\zeta(t)$ by

$$\zeta(t) = A(t)^{-1}x(t) = e^{-tQ}x(t). \tag{4.3.14}$$

Then if $s_1 < t_1 < s_2 < t_2$

$$E\{\zeta(t_1) - \zeta(s_1)\} = 0, \qquad E\{[\zeta(t_2) - \zeta(s_2)]\cdot[\zeta(t_1) - \zeta(s_1)]\} = 0 \tag{4.3.15}$$

and

$$D(t) = E[\zeta(t) - \zeta(0)]\cdot[\zeta(t) - \zeta(0)]\} = e^{-tQ}R(0)e^{-tQ^*} - R(0). \tag{4.3.16}$$

Hence the $\{\zeta(t)\}$ determine a ζ-process, with dispersion matrix given by (4.3.16). The derivative $D'(t)$ is easily evaluated:

$$D'(t) = e^{-tQ}[-R(0)Q^* - QR(0)]e^{-tQ^*}. \tag{4.3.17}$$

Since $D'(t)$ is symmetric and non-negative definite, the bracket also has this property, and there is a non-singular matrix S_1 such that

$$S_1[-QR(0) - R(0)Q^*]S_1^* = U, \tag{4.3.18}$$

where U is in diagonal form, with only 0's and 1's in the main diagonal. Then the integral

$$\int_0^t S_1 e^{sQ}\, d\zeta(s) \tag{4.3.19}$$

defines a ζ-process with dispersion matrix tU. There is therefore a ξ-process with variables $\{\xi(t)\}$ such that

$$U\xi(t) = \int_0^t S_1 e^{sQ}\, d\zeta(s). \tag{4.3.20}$$

This equation can be solved for $\zeta(t)$ and $x(t)$:

$$x(t) = e^{tQ}\zeta(t) = e^{tQ}\int_0^t e^{-sQ}S_2 U\, d\zeta(s) + e^{tQ}x(0) \tag{4.3.21}$$

where $S_2 = S_1^{-1}$. The matrix S_2U can be written in the polar form $S0$ where S is symmetric and non-negative definite and 0 is orthogonal. This S is the S of (4.3.2) etc.

The remainder of the proof follows closely the proof of Theorem 3.6 and will be omitted.

An important class of t.h.G.M. processes which arises frequently in physical applications is obtained in the following way. Let $\{\xi(t)\}$ be the variables of a one-dimensional ξ-process. Consider the formal equation

$$\frac{d^N y(t)}{dt^N} - a_1 \frac{d^{N-1} y(t)}{dt^{N-1}} - \cdots - a_N y(t) = c\xi'(t), \tag{4.4.1}$$

where $a_1, \cdots, a_N, c$ are constants. This equation cannot be considered precise as it stands, since $\xi'(t)$ does not exist. The problem can however be reformulated as follows: find a $y(t)$ process, where $y', \cdots, y^{(N-1)}$ are supposed to exist, satisfying the equation

$$\int_a^b f(t)\, dy^{(N-1)}(t) - a_1 \int_a^b f(t)\, dy^{(N-2)}(t) - \cdots - a_N \int_a^b f(t) y(t)\, dt = c \int_a^b f(t)\, d\xi(t) \tag{4.4.2}$$

with probability 1, for each continuous function $f(t)$ and each pair of numbers a, b. The formal integrals are defined as the limit in the mean of the usual sums.[20] The integral on the right has already been discussed. With this interpretation, equations involving ξ' can be treated in the usual way, and this will be done in the following without further comment. The formal solution of (4.4.2) is well known. Let $\lambda_1, \cdots, \lambda_N$ be the roots of the equation

$$\lambda^N - a_1\lambda^{N-1} - \cdots - a_N = 0 \tag{4.4.3}$$

and suppose that these roots are distinct, and have negative real parts. Let Δ_{jk} be the cofactor of λ_j^{k-1} in the determinant

$$\delta = \begin{vmatrix} 1 & \cdot & \cdot & \cdot & 1 \\ \lambda_1 & \lambda_2 & \cdot & \cdot & \lambda_N \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \lambda_1^{N-1} & \lambda_2^{N-1} & \cdot & \cdot & \lambda_N^{N-1} \end{vmatrix}. \tag{4.4.4}$$

Then the general solution of (4.4.1), that is to say of (4.4.2), is

$$y(t) = \frac{c}{\delta} \int_0^t \sum_{j=1}^N \Delta_{Nj} e^{\lambda_j (t-s)}\, d\xi(s) + \frac{1}{\delta} \sum_{j,k=1}^N \Delta_{kj} e^{\lambda_j t} y^{(k-1)}(0). \tag{4.4.5}$$

Since the integrand and its first $N-1$ derivatives vanish when $s = t$, $y', \cdots, y^{(N-1)}$ as defined by (4.4.5) exist, but $y^{(N)}(t)$ does not exist, because $\xi'(t)$ in (4.4.1)

[20] For a full discussion in the case $N = 1$ cf. Doob, *Annals of Math.*, Vol. 43 (1942), pp. 358-61.

does not exist. The $y(t)$ process is a t.h.G. process if $y(0), \cdots, y^{(N-1)}(0)$ are chosen properly. This can be seen from the solution

$$y(t) = \frac{c}{\delta} \int_{-\infty}^{t} \sum_{j=1}^{N} \Delta_{Nj} e^{\lambda_j (t-s)}\, d\xi(s). \tag{4.4.6}$$

In fact this is the only solution defining a t.h.G. process. To prove this, rewrite (4.4.5) in the form

$$y(t) = \frac{c}{\delta} \sum_{j=1}^{N} \Delta_{Nj} \int_{t-\tau}^{t} e^{\lambda_j (t-s)}\, d\xi(s) + \frac{1}{\delta} \sum_{j,k=1}^{N} \Delta_{kj} e^{\lambda_j (t-\tau)} y^{(k-1)}(\tau). \tag{4.4.5'}$$

If the $y(t)$ process is a t.h.G. process, (4.4.5′) becomes (4.4.6) when $\tau \to -\infty$. Thus there is a unique stationary solution to (4.4.1) and, by (4.4.5), every solution tends to this solution in the long run. The stationary solution (4.4.6) has the property that $y(t)$ is written in terms of $\xi(s)$ for $s \leqq t$. Then in (4.4.5) the integral is independent of the terms involving the initial conditions. In other words

$$E\{y(s),\ s \leqq 0;\ y(t)\} = \frac{1}{\delta} \sum_{j,k=1}^{N} \Delta_{kj} e^{\lambda_j t} y^{(k-1)}(0). \tag{4.4.7}$$

Hence the variables $y(t), y'(t), \cdots, y^{(N-1)}(t)$ define an N-dimensional t.h.G.M. process. The transition matrix function $A(t)$, and the matrices Q, S, T of Theorem 4.3 are easily calculated.

$$A(t): (\Delta_{kj} \lambda_j^{k-1} e^{\lambda_j t}),$$

$$Q: \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & 1 \\ a_N & \cdot & \cdot & \cdot & a_1 \end{pmatrix} \tag{4.4.8}$$

$$T = 0.$$

The necessary changes to be made if the λ_j are not distinct are well known. The case $c = 0$ will be treated below, when the problem will be reconsidered from another point of view. In all cases the solution of (4.4.1) leads to an N-dimensional t.h.G.M. process.[21]

As a simple example, consider a torsion pendulum, suspended in a sealed container. The only turning forces acting on the pendulum are the molecular shocks of the surrounding gas, and the restoring torque. The equation of motion is

$$I \frac{d^2 y(t)}{dt^2} + \alpha_1 \frac{dy(t)}{dt} + \alpha_2 y(t) = X(t), \tag{4.4.9}$$

[21] According to a letter from Uhlenbeck, the differential equation (4.4.1) was solved; from a somewhat different point of view, by Miss Ming Chen Wang, in a thesis written in 1941 which is unfortunately inaccessible to me at the moment.

where y is the angular displacement measured from the equilibrium position, I is the moment of inertia, α_2 is the torque coefficient of the suspension, and the molecular force is resolved into a systematic Stokes term $\alpha_1 y'$ and a remainder X. The remainder term $X(t)$ defines a stationary process which to a first approximation is "purely random." In the present context "purely random" means that if $t_1 < \cdots < t_\nu$, $X(t_1), \cdots, X(t_\nu)$ are mutually independent. This is precisely the property the derived process of a ξ-process would have, if $\xi'(t)$ existed. Unfortunately it has already been noted that $\xi'(t)$ does not exist, since the difference quotient $[\xi(t + h) - \xi(t)]/h$ is unbounded as $h \to 0$. It has already been seen, however, that (4.4.9) can be given a meaning with $X(t)$ identified with $c\xi'(t)$ even though $\xi'(t)$ does not exist, and it has been seen that the solution approaches a steady state. It may still be a disappointment to some that the solution $y(t)$ has a first derivative $y'(t)$ but that $y''(t)$ does not exist: there is an angular velocity but not an angular acceleration! This unhappy circumstance can either be blamed on the physical world, or on the mathematical approximation to the physical world, depending on the point of view. The corresponding electrical picture is the following. There are spontaneous currents in any electrical circuit, due to the thermal motion of the electrons. This is known as the Johnson effect. In a simple closed circuit, consisting of an inductance L, a resistance, R, and a capacitance C in series, the current equation can be written in the form

$$L\frac{d^2y(t)}{dt^2} + R\frac{dy(t)}{dt} + \frac{y(t)}{C} = X(t), \tag{4.4.10}$$

where y is the charge on the condenser and $X(t)$ represents a fictitious voltage set up by the motion of the electrons. The $X(t)$ is identified with $c\xi'(t)$ as before. In this case there is a current $\frac{dy}{dt}$, but the current function has no derivative. In these applications, the physical justification for the Gaussian character of the ξ-distribution lies in the Gaussian character of the Maxwell distribution of elementary particle velocities. The known mean particle kinetic energy determines the constant c in (4.4.1). The more complicated mechanical or electrical systems will lead to equations of higher order than 2, or systems of equations. For example the usual current equations of a net or resistances capacitances and inductances lead to a system of say ν second order equations of type (4.4.10), and the corresponding pairs y, y' form a 2ν-dimensional t.h.G.M. process.[22]

The processes defined by linear differential equations of the type (4.4.1) are the analogues of the t.h.G.M$_N$. processes in the discrete parameter case. Instead of defining these solutions of (4.4.1) as the t.h.G.M$_N$. processes, however, we shall use a definition closer to the definition in the discrete parameter case. A

[22] Further discussion and references to papers by physicists on this subject will be found in Doob, *Annals of Math.*, Vol. 43 (1942), pp. 351–69.

one-dimensional t.h.G. process with variables $\{y(t)\}$ will be called a t.h.G.M$_N$. process if the derivatives $y'(t), \cdots, y^{(N-1)}(t)$ exist, and if whenever $s < t$,

$$(4.5.1)\quad E\{y(\tau),\ \tau \leqq s;\ y(t)\} = E\{y(s),\ y'(s), \cdots, y^{(N-1)}(s);\ y(t)\}.$$

If $N = 1$, the process is a t.h.G.M. process. The right hand side of (4.5.1) is a linear combination of the variables $y(s), \cdots, y^{(N-1)}(s)$. The variables $\{y(t)\}$ thus satisfy an equation of the form

$$(4.5.2)\quad y(t) - a_1(t-s)y(s) - \cdots - a_N(t-s)y^{(N-1)}(s) = \eta(s, t)$$

where $\eta(s, t)$ is independent of the variables $\{y(\tau)\}$ for $\tau \leqq s$. Define the variables $\{x(t)\}$ of an N-dimensional t.h.G. process by

$$(4.5.3)\quad \begin{aligned} x_1(t) &= y(t) \\ x_j(t) &= y^{(j)}(t), \qquad j = 1, \cdots, N-1. \end{aligned}$$

If this process is degenerate, there is a relation of the form

$$(4.5.4)\quad c_0 y(s) + c_1 y'(s) + \cdots + c_{N-1} y^{N-1}(s) = 0, \qquad \sum_{j=0}^{N-1} |c_j| > 0.$$

It can be assumed that $c_{N-1} \neq 0$, (differentiating (4.5.4) to get a term in $y^{(N-1)}(t)$ if there is none originally). Then $y^{(N-1)}(s)$ can be eliminated in (4.5.2), to get a relation of the same type with N replaced by $N-1$. Hence the process is non-degenerate if N is the minimum index for which (4.5.1) is true. It will now be proved that the $x(t)$ process is a t.h.G.M. process. It can be assumed to be non-degenerate. Using (4.5.1),

$$(4.5.5)\quad E\{x(\tau),\ \tau \leqq s;\ x_1(t)\} = E\{y(\tau),\ \tau \leqq s;\ y(t)\} = E\{x(s);\ x_1(t)\}.$$

It must also be shown that

$$(4.5.6)\quad E\{x(\tau),\ \tau \leqq s;\ x_j(t)\} = E\{x(s);\ x_j(t)\} \qquad j = 2, \cdots, N.$$

This will be shown by justifying the taking of derivatives in (4.5.5). It will be sufficient to prove (4.5.6) when $j = 2$. Using (4.5.1),

$$(4.5.7)\quad E\left\{x(\tau),\ \tau \leqq s;\ \frac{y(t+h) - y(t)}{h}\right\} = E\left\{x(s);\ \frac{y(t+h) - y(t)}{h}\right\}.$$

The right hand side is a linear combination of $x_1(s), \cdots, x_N(s)$ whose coefficients are continuous in h, $h \geqq 0$, since the correlation function of the $y(t)$ process is continuous. Hence the right hand side converges to

$$E\{x(s);\ y'(t)\} = E\{x(s);\ x_2(t)\}$$

when $h \to 0$. Since the difference

$$x(t) - E\left\{x(s);\ \frac{y(t+h) - y(t)}{h}\right\}$$

is uncorrelated with $x(\tau)$ if $\tau \leqq s$, the same is true of its limit as $h \to 0$. This means that (4.5.6) is true when $j = 2$, as was to be shown. Conversely if $\{y(t)\}$ are the variables of a one-dimensional t.h.G. process, if $y'(t), \cdots, y^{(N-1)}(t)$ exist, and if the $x(t)$ process defined by (4.5.3) is a t.h.G.M. process, the $y(t)$ process is obviously a t.h.G.M$_N$. process. The transition matrix function $A(t)$ and the matrices Q, S, T of Theorem 4.3 are easily calculated. Suppose that the $x(t)$ process is non-degenerate. Since $y^{(i-1)}(t)$ is given by

$$\begin{aligned} y^{(i-1)}(t) = x_i(t) &= \int_{-\infty}^{t} \sum_{j=1}^{N} [e^{(t-s)Q} S]_{ij}\, d\xi_j(s) + \sum_{j=1}^{N} [e^{tQ} T]_{ij} \xi_j \\ &= \int_{0}^{t} \sum_{j=1}^{N} [e^{(t-s)Q} S]_{ij}\, d\xi_j(s) + \sum_{j=1}^{N} (e^{tQ})_{ij} x_j(0) \end{aligned} \tag{4.5.8}$$

and since $x_i'(t)$ exists if $i < N$, it follows that the integrand must vanish when $s = t$:

$$(S)_{ij} = 0, \qquad i = 1, \cdots, N-1, \qquad j = 1, \cdots, N. \tag{4.5.9}$$

Since S is symmetric and non-negative definite, S must have the form

$$S: \begin{pmatrix} 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & c \end{pmatrix}, \qquad c \geqq 0. \tag{4.5.10}$$

The fact that $x_i'(t) = x_{i+1}(t)$ means that

$$\sum_{j=1}^{N} (e^{tQ} Q)_{ij} x_j(0) = \sum_{j=1}^{N} (e^{tQ})_{i+1\,j} x_j(0), \qquad j = 1, \cdots, N \tag{4.5.11}$$

or, since the $x(t)$ process is non-degenerate,

$$(e^{tQ} Q)_{ij} = (e^{tQ})_{i+1\,j} \qquad \begin{matrix} i = 1, \cdots, N-1 \\ j = 1, \cdots, N. \end{matrix} \tag{4.5.12}$$

Hence $(t \to 0)$ Q has the form

$$Q: \begin{pmatrix} 0 & 1 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & 1 \\ a_N & \cdot & \cdot & \cdot & \cdot & a_1 \end{pmatrix}. \tag{4.5.13}$$

Conversely if there is an N-dimensional non-degenerate t.h.G.M. process with transition matrix function e^{tQ} where Q is given by (4.5.13) and dispersion matrix S given by (4.5.10),

$$x_i'(t) = x_{i+1}(t), \qquad i = 1, \cdots, N-1,$$

and the $x_1(t)$ process is a t.h.G.M$_N$. process.

Case 1. $S = 0$ (*deterministic case*). In this case the $x(t)$ process is deterministic:

$$x(t) = e^{tQ}T\xi. \tag{4.5.14}$$

Since Q satisfies its characteristic equation

$$\alpha^N - a_1\alpha^{N-1} - \cdots - a_N = 0 \tag{4.5.15}$$

it follows that

$$x^{(N)}(t) - a_1x^{(N-1)}(t) - \cdots - a_Nx(t) = 0, \tag{4.5.16}$$

$$y^{(N)}(t) - a_1y^{(N-1)}(t) - \cdots - a_Ny(t) = 0. \tag{4.5.17}$$

The roots of (4.5.15) are simple roots, and are all pure imaginary, according to Theorem 4.3. It follows that

$$y(t) = \sum_j (\eta_j \cos t\theta_j + \zeta_j \sin t\theta_j) \tag{4.5.18}$$

where the η_j and ζ_j are one-dimensional Gaussian variables, and $\{i\theta_j\}$ are the distinct roots of (4.5.15).

Case 2. $S \neq 0$ (*non-deterministic case*). In this case it will now be shown that the $x(t)$ process has no deterministic factor, that is that the roots of (4.5.15) all have negative real parts. In fact let β be a root of (4.5.15), corresponding to the characteristic vector z of Q^*:

$$\begin{aligned} z &= (a_N\beta^{N-1}, a_N\beta^{N-2} + a_{N-1}\beta^{N-1}, \cdots, a_N + a_{N-1}\beta + \cdots + a_1\beta^{N-1}) \\ &= (\cdots, \beta^N). \end{aligned} \tag{4.5.19}$$

Then using (4.3.5)

$$\begin{aligned} 0 < c^2 \mid \beta \mid^{2N} = (S^2z, z) &= -(R(0)Q^*z, z) - (QR(0)z, z) \\ &= -\lambda(R(0)z, z) - \bar{\lambda}(R(0)z, z) \\ &= -(\lambda + \bar{\lambda})(R(0)z, z). \end{aligned} \tag{4.5.20}$$

Hence $\lambda + \bar{\lambda}$ is real and negative: λ has a negative real part. In this non-deterministic case, therefore, the $x(t)$ process can have no deterministic factor. The matrix T is the null matrix, and (4.3.2) becomes

$$x(t) = \int_{-\infty}^{t} e^{(t-s)Q} S \, d\xi(s) \tag{4.5.21}$$

which leads to

$$y(t) = c \int_{-\infty}^{t} [e^{(t-s)Q}]_{1N} \, d\xi_N(s). \tag{4.5.22}$$

Moreover

$$
\begin{aligned}
y'(t) &= c\int_{-\infty}^{t} [Qe^{(t-s)Q}]_{1N}\, d\xi_N(s) \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
y^{(N-1)}(t) &= c\int_{-\infty}^{t} [Q^{N-1}e^{(t-s)Q}]_{1N}\, d\xi_N(s).
\end{aligned} \tag{4.5.23}
$$

Since Q satisfies its characteristic equation (4.5.15),

$$
\int_{-\infty}^{t} [Q^N e^{(t-s)Q}]_{1N}\, d\xi_N(s) - \sum_{j=1}^{N} a_j \int_{-\infty}^{t} [Q^{N-j}e^{(t-s)Q}]_{1N}\, d\xi_N(s) = 0. \tag{4.5.24}
$$

In other words

$$
c\int_{-\infty}^{t} [Q^N e^{(t-s)Q}]_{1N}\, d\xi_N(s) - a_1 y^{(N-1)}(t) - \cdots - a_N y(t) = 0. \tag{4.5.25}
$$

Now formally, if $\xi_N'(t)$ existed, the last equation in (4.5.23) could be differentiated to give

$$
y^{(N)}(t) = c[Q^{N-1}]_{1N}\,\xi_N'(t) + c\int_{-\infty}^{t} [Q^N e^{(t-s)Q}]_{1N}\, d\xi_N(s) \tag{4.5.26}
$$

and (4.5.25) would become

$$
y^{(N)}(t) - a_1 y^{(N-1)}(t) - \cdots - a_N y(t) = c\xi_N'(t). \tag{4.5.27}
$$

(We are using the fact that $(Q^{N-1})_{1N} = 1$.) Thus the t.h.G.M$_N$. processes satisfy the formal differential equation (4.5.27) already discussed above from another point of view. Equation (4.4.2) is readily justified.

THEOREM 4.6. (i) *Let $\{x(t)\}$ be the variables determining a t.h.G.M. process. Then considered as functions of t, the $x(t)$ are continuous with probability* 1. *Let $\{y(t)\} = \{x_r(t)\}$ be the variables of a coordinate process.*

(ii) *If $y'(t)$ exists, it is a linear combination of coordinate functions*:

$$
y'(t) = \sum_{j=1}^{N} c_j x_j(t).
$$

(iii) *If $y'(t), \cdots, y^{(N-1)}(t)$ exist, $y(t)$ satisfies a generalized differential equation* (4.4.1), *that is the $y(t)$ process is a t.h.G.M$_N$. process.*

(iv) *If $y'(t), \cdots, y^{(N)}(t)$ exist, $y(t)$ has derivatives of all orders. The $y(t)$ process is a t.h.G.M$_N$. process (deterministic case) and $y(t)$ therefore satisfies an Nth order homogeneous differential equation* (4.5.17).

(v) *If $x_1'(t), \cdots, x_N'(t)$ exist, that is if $x'(t)$ exists, the $x(t)$ process is deterministic and the coordinate functions have derivatives of all orders.*

Proof of (*i*). It has already been shown that the $\{\zeta(t)\}$ determined by (4.3.14) determine a ζ-process, and the dispersion matrix function $D(t)$ of the ζ-process, given by (4.3.16), is certainly continuous. Hence, by Theorem 4.1, the $\{\zeta(t)\}$, and therefore the $\{x(t)\}$ are continuous in t, with probability 1.

Proof of (ii). If $x_r'(t)$ exists, the rth row of S in (4.3.2) must vanish, and $x_r'(t)$ is given by the rth coordinate of

$$\int_{-\infty}^{t} Qe^{(t-s)Q}\, S\, d\xi(s) + Qe^{tQ}\, x(0) = Qx(t). \tag{4.6.1}$$

Hence

$$x_r'(t) = \sum_{j=1}^{N} (Q)_{rj}\, x_j(t). \tag{4.6.2}$$

Proof of (iii). Suppose that $x_r'(t), \cdots, x_r^{(N-1)}(t)$ exist. Then (r is fixed in the following equations)

$$\begin{aligned}
x_r(t) &= \int_0^t \sum_{j=1}^{N} [e^{(t-s)Q}\, S]_{rj}\, d\xi_j(s) + \sum_{j=1}^{N} [e^{tQ}]_{rj}\, x_j(0), \\
x_r'(t) &= \int_0^t \sum_{j=1}^{N} [e^{(t-s)Q}\, QS]_{rj}\, d\xi_j(s) + \sum_{j=1}^{N} [e^{tQ}\, Q]_{rj}\, x_j(0), \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
x_r^{(N-1)}(t) &= \int_0^t \sum_{j=1}^{N} [e^{(t-s)Q}\, Q^{N-1} S]_{rj}\, d\xi_j(s) + \sum_{j=1}^{N} [e^{tQ}\, Q^{N-1}]_{rj}\, x_j(0)
\end{aligned} \tag{4.6.3}$$

and (in order that the derivatives can exist)

$$\begin{aligned}
S_{rj} &= 0 \\
(QS)_{rj} &= 0 \\
&\cdots\cdots\cdots \\
(Q^{N-2} S)_{rj} &= 0,
\end{aligned} \qquad j = 1, \cdots, N. \tag{4.6.4}$$

Since Q satisfies its characteristic equation, say (4.5.15),

$$\int_{-\infty}^{t} [e^{(t-s)Q}\, Q^N\, S]\, d\xi(s) - \sum_{j=1}^{N} a_j \int_{-\infty}^{t} [Q^{N-j} e^{(t-s)Q}]\, d\xi(s) = 0. \tag{4.6.5}$$

This vector equation can be written (using only the rth coordinate) in the form

$$\int_{-\infty}^{t} \sum_{j=1}^{N} [e^{(t-s)Q}\, Q^N\, S]_{rj}\, d\xi_j(s) - a_1 x_r^{(N-1)}(t) - \cdots - a_N x_r(t) = 0. \tag{4.6.6}$$

If $\xi'(t)$ existed, the last equation in (4.6.3) could be differentiated to give

$$x_r^{(N)}(t) = \int_{-\infty}^{t} \sum_{j=1}^{N} [e^{(t-s)Q}\, Q^N\, S]_{rj}\, d\xi_j(s) + \sum_{j=1}^{N} [Q^{N-1}\, S]_{rj}\, \xi_j'(t) \tag{4.6.7}$$

and (4.6.7) would then become

$$x_r^{(N)}(t) - a_1 x_r^{(N-1)}(t) - \cdots - a_N x_r(t) = \sum_{j=1}^{N} (Q^{N-1} S)_{rj}\, \xi_j'(t). \tag{4.6.8}$$

Now the process with variables

$$\left\{\frac{1}{c} \sum_{j=1}^{N} (Q^{N-1} S)_{rj}\, \xi_j(t)\right\} \tag{4.6.9}$$

is a ξ-process, if c is chosen properly, unless the parenthesis in (4.6.9) vanishes for all j. In either case (iii) is proved.

Proof of (iv). If in (iii), $x_r^{(N)}(t)$ exists, (4.6.3) can be augmented to include

$$x_r^{(N)}(t) = \int_{-\infty}^{t} \sum_{j=1}^{N} [e^{(t-s)Q} Q^N S]_{rj}\, d\xi_j(s) + \sum_{j=1}^{N} [e^{tQ} Q^N]x_j(0) \tag{4.6.3'}$$

and (4.6.4) now includes

$$(Q^{N-1}S)_{rj} = 0, \qquad j = 1, \cdots, N. \tag{4.6.4'}$$

In this case the last term in (4.6.7) vanishes and (4.6.8), with zero on the right hand side, is strictly true.

Proof of (v). If $x_1'(t), \cdots, x_N'(t)$ exist, S must vanish and (4.6.3) yields

$$x(t) = e^{tQ}x(0), \qquad x^{(\nu)}(t) = Q^\nu x(t). \tag{4.6.10}$$

Thus the $x(t)$ process is deterministic and $x(t)$ has derivatives of all orders.

THEOREM 4.7. *Let $\{x(t)\}$ be the variables of a one-dimensional t.h.G. process. The process is a component process of an N-dimensional t.h.G.M. process if and only if the chance variables*

$$x(0), \{E\{x(s), s \leqq 0; x(t)\}\} \qquad 0 < t < \infty \tag{4.7.1}$$

are linearly dependent on N variables.

Suppose that the $x(t)$ process is a component process of an N-dimensional t.h.G.M. $y(t)$ process: $x(t) = y_r(t)$, and let $A(t)$ be the transition matrix function of the $y(t)$ process. Then if $\epsilon > 0$ and if n is any integer, the difference

$$y[(n+1)\epsilon] - A(\epsilon)y(n\epsilon)$$

is independent of every $y(s)$ with $s \leqq n\epsilon$, and therefore independent of every $y(m\epsilon)$ with $m \leqq n$. Hence the $y(n\epsilon)$ process is a t.h.G.M. process (discrete parameter case). Equation (3.8.5) becomes, in this case, if $n = 0$,

$$x[(N+\nu)\epsilon] - a_1^{(\nu)} x[(N-1)\epsilon] - \cdots - a_N^{(\nu)} x(0) = \eta_1^{(\nu)}(N+\nu) \tag{4.7.2}$$

where $\eta_1^{(\nu)}(N+\nu)$ is not merely independent of the variables $\cdots, x(-\epsilon), x(0)$, but is even independent of every $x(s)$ with $s \leqq 0$. It then follows, applying the operator $E\{x(s), s \leqq 0; \cdot\}$ to both sides of (4.7.2), that the variables in (4.7.1) are linearly dependent on N variables if t is restricted to be a multiple of ϵ. Allowing ϵ to run through the values

$$\left\{\frac{1}{m!}\right\}, \qquad m = 1, 2, \cdots$$

it follows that the statement of the theorem is true if t is restricted to be rational. The proof will be complete when it is shown that the subject[23] variables for rational t are dense in the whole class in the sense that for any t, the expectation

$$\delta = E\{[E\{x(s), s \leqq 0; x(t')\} - E\{x(s), s \leqq 0; x(t)\}]^2\} \tag{4.7.3}$$

[23] Courtesy of U. S. Navy.

converges to 0 when $t' \to t$. In fact, using the Schwarz inequality

$$\begin{aligned} \delta &= E\{[E\{x(s), s \leqq 0; x(t') - x(t)\}]^2\} \\ &\leqq E\{E\{x(s), s \leqq 0; [x(t') - x(t)]^2\}\} = E\{[x(t') - x(t)]^2\} \end{aligned} \tag{4.7.4}$$

and the basic continuity hypothesis (1.3.1) imposed on continuous processes is precisely that the last expectation converges to 0 when $t' \to t$.

Conversely suppose that the chance variables (4.7.1) are linearly dependent on N variables. It can be supposed that $x(0)$ is one of these N. Let the others be those for which $t = t_2, \cdots, t_N$, and define $y_1(t), \cdots, y_N(t)$ by

$$\begin{aligned} y_1(t) &= x(t) \\ y_j(t) &= E\{x(s), s \leqq t; x(t + t_j) \qquad j = 2, \cdots, N. \end{aligned} \tag{4.7.5}$$

The $y(t)$ process is obviously an N-dimensional t.h.G. process. Moreover

$$\begin{aligned} E\{y(s), s \leqq 0; y_j(t) &= E\{x(s), s \leqq 0; y_j(t)\} \\ &= E\{x(s), s \leqq 0; x(t + t_j)\} \qquad j = 1, \cdots, N \end{aligned} \tag{4.7.6}$$

(where t_1 is defined as 0). Since the right side is by hypothesis, for each j, a linear combination of $y_1(0), \cdots, y_N(0)$, the $y(t)$ process is a t.h.G.M. process, and the $x(t)$ process is a component process, as was to be shown.

A detailed examination will now be made of t.h.G.M$_N$. processes, and of the more general class of component processes of t.h.G.M. processes. The following theorem will be useful.

THEOREM 4.8. *Let $\{x(t)\}$ be the variables determining a t.h.G. continuous parameter process. The process is a component process of an N-dimensional t.h.G.M. process if and only if for each $\epsilon > 0$ the discrete parameter process with variables $\{x(n\epsilon)\}$ is a component process of an N-dimensional t.h.G.M. process.*

If the $x(t)$ process is a component process of an N-dimensional t.h.G.M. $y(t)$ process, the $x(n\epsilon)$ process is a component process of the N-dimensional t.h.G.M. $y(n\epsilon)$ process. Conversely suppose that the $x(n\epsilon)$ process is a component process of an N-dimensional t.h.G.M. process (which may depend on ϵ) for every $\epsilon > 0$. It follows that for each $\epsilon > 0$ the chance variables

$$E\{\cdots, x(-\epsilon), x(0); x(n\epsilon)\}, \qquad n = 0, 1, \cdots \tag{4.8.1}$$

are linearly dependent on N of their number. Hence the same is true of the following chance variables, if ν, m are fixed and $\nu > m$:

$$E\{\cdots, x(-1/\nu!), x(0); x(n/m\} \qquad n = 0, 1, \cdots. \tag{4.8.2}$$

According to the Corollary to Theorem 1.2, when $\nu \to \infty$ the conditional expectations in (4.8.2) converge to

$$E\{x(s), s \leqq 0, s \text{ rational}; x(n/m)\} \qquad n = 0, 1, \cdots. \tag{4.8.3}$$

Hence the chance variables (t rational)

$$E\{x(s),\ s \leqq 0,\ s \text{ rational};\ x(t)\} = E\{x(s),\ s \leqq 0;\ x(t)\}^{24}, \tag{4.8.4}$$
$$0 < t < \infty,$$

are linearly dependent on N of their number. As in the proof of Theorem 4.7 it follows that the same is true if t runs through all positive real numbers, and according to Theorem 4.7, the $x(t)$ process is therefore a component process of an N-dimensional t.h.G.M. process.

THEOREM 4.9. *Let $\{x(t)\}$ be the variables of a one-dimensional continuous parameter t.h.G. process. The process is a component process of a finite-dimensional t.h.G.M. process if and only if the complex spectral function of the process is the sum of the integral of the square of the absolute value of a rational function of λ and of a monotone non-decreasing function increasing only in a finite number of jumps.*[25] *Specifically:*

(i) *The process is a component process of an N-dimensional t.h.G.M. process if and only if the complex spectral function has the form*

$$G(\lambda) = \int_{-\infty}^{\lambda} \frac{|\beta_0(i\lambda)^{N-1} + \cdots + \beta_{N-1}|^2}{|(i\lambda)^N + \alpha_1(i\lambda)^{N-1} + \cdots + \alpha_N|^2}\, d\lambda + \hat{G}(\lambda) \tag{4.9.1}$$

where

(a) *$\hat{G}(\lambda)$ is a monotone non-decreasing function satisfying* (1.3.3) *and increasing only in jumps, at no more than N points.*

(b) *the denominator of the integrand vanishes at every discontinuity of $\hat{G}(\lambda)$, and the numerator vanishes at every zero of the denominator, to at least the same order;*

(c) *the coefficients in the integrand are real, and the roots of the λ polynomials are all on the real axis or in the upper half plane.*

The integral vanishes identically if and only if the $x(n)$ process is a component process of an N-dimensional deterministic process, and $\hat{G}(\lambda)$ vanishes identically if and only if the variables $\{x(t)\}$ vanish identically or the $x(t)$ process is a component process of an N-dimensional t.h.G.M. process with no deterministic factor.

(ii) *The process is a t.h.G.M$_N$. process, in the deterministic case, if and only if the complex spectral function $G(\lambda) = \hat{G}(\lambda)$ is a function increasing only in jumps, at no more than N points; non-deterministic case if and only if the complex spectral function has the form*

$$G(\lambda) = \int_{-\infty}^{\lambda} \frac{c\, d\lambda}{|(i\lambda)^N + \cdots + \alpha_N|^2}. \tag{4.9.2}$$

---

[24] The equality (4.8.4) is proved as follows. Let t be fixed, and let x be the chance variable on the left. Then $x(t) - x$ has mean 0 and is uncorrelated with every $x(s)$ with $s \leqq 0$ and rational. It follows at once from the continuity of hypothesis (1.3.1) that then $x(t) - x$ is uncorrelated with every $x(s)$ with $s \leqq 0$: it follows that (4.8.4) is true.

[25] It is easily seen that the first term of the two can also be described simply as the integral of a rational function of λ, which is non-negative for real λ and is integrable and an even function, like all complex spectral density functions.

Proof of (*i*). Suppose that the $x(t)$ process is a one-dimensional component process of an N-dimensional t.h.G.M. $y(t)$ process, $x(t) = y_1(t)$. It is no restriction to assume that the $y(t)$ process is non-singular. Then the correlation function of the $y(t)$ process is given by

$$\begin{aligned} R_y(t) &= R_y(0)e^{tQ^*} \qquad t \geqq 0 \\ R_y(t) &= e^{-tQ}R_y(0) \qquad t \leqq 0, \end{aligned} \tag{4.9.3}$$

where Q is uniquely determined and

$$\begin{aligned} G(\lambda) - G(0) &= \left[\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{it\lambda}-1}{it}R_y(t)\,dt\right]_{11} \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{it\lambda}-1}{it}[R_y(t)]_{11}\,dt, \end{aligned} \tag{4.9.4}$$

at the points of continuity of $G(\lambda)$.

The correlation function $R_y(t)$ has derivatives of all orders for $t > 0$:

$$\begin{aligned} R_y^{(\nu)}(t) &= R_y(0)Q^{*\nu}e^{tQ^*} \qquad t > 0 \\ &= (-1)^\nu Q^\nu e^{-tQ}R_y(0) \qquad t < 0. \end{aligned} \tag{4.9.5}$$

Suppose first that the $y(t)$ process has no deterministic factor, in other words that it is non-degenerate and of type M. Then the characteristic values of Q have negative real parts and $R(t) \to 0$ exponentially when $|t| \to \infty$. Hence $G(\lambda)$ has a continuous derivative $G'(\lambda)$:

$$G'(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{\infty}e^{it\lambda}[R_y(t)]_{11}\,dt. \tag{4.9.6}$$

Integrating by parts,

$$\begin{aligned} G'(\lambda) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{it\lambda}}{i\lambda}[R_y'(t)]_{11}\,dt \\ &= \frac{R_y'(0+) - R_y'(0-)}{2\pi(i\lambda)^2} + \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{it\lambda}}{(i\lambda)^2}R_y''(t)\,dt \\ &= \frac{R_y'(0+) - R_y'(0-)}{2\pi(i\lambda)^2} - \frac{R_y''(0+) - R_y''(0-)}{2\pi(i\lambda)^3} - \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{it\lambda}}{(i\lambda)^3}R_y'''(t), \\ &\quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{4.9.7}$$

Since Q satisfies its characteristic equation

$$\alpha^N - a_1\alpha^{N-1} - \cdots - a_N = 0, \tag{4.9.8}$$

it follows that

$$\begin{aligned} R_y^{(N)}(t) - a_1R_y^{(N-1)}(t) - \cdots - a_NR_y(t) &= 0 \qquad t > 0 \\ R_y^{(N)}(t) + a_1R_y^{(N-1)}(t) - \cdots + (-1)^{N-1}R_y(t) &= 0 \qquad t < 0 \end{aligned} \tag{4.9.9}$$

and therefore if U is the operator $\frac{d}{dt}$

$$[U^N - a_1U^{N-1} - \cdots - a_NU^0][U^N + a_1U^{N-1} - \cdots + (-1)^{N-1}U^0]R_y(t)=0, \quad t \neq 0. \tag{4.9.10}$$

Applying (4.9.10) to (4.9.6)

$$[(i\lambda)^N - a_1(i\lambda)^{N-1} - \cdots - a_N][(i\lambda)^N + a_1(i\lambda)^{N-1} + \cdots + (-1)^{N-1}a_N]\,G'(\lambda) = |(i\lambda)^N - a_1(i\lambda)^{N-1} - \cdots - a_N|^2 G'(\lambda) = P(i\lambda) \tag{4.9.11}$$

where $P(i\lambda)$ is a polynomial of degree $2n - 2$. Since $P(i\lambda)$ is real and non-negative, when λ is real, the roots on the real axis are of even multiplicity and those off the axis are symmetric in the axis. Moreover $P(i\lambda)$ is even, since the left side of (4.9.11) is even. It follows easily that $P(i\lambda)$ can be written in the form

$$P(i\lambda) = \left|\sum_{j=0}^{N-1} \beta_j(i\lambda)^{N-j-1}\right|^2 \tag{4.9.12}$$

where the roots of the β polynomial are all on or to the left of the imaginary axis. Finally

$$G'(\lambda) = \frac{|\beta_0(i\lambda)^{N-1} + \cdots + \beta_{N-1}|^2}{|(i\lambda)^N - \cdots - a_N|^2}. \tag{4.9.13}$$

The denominator polynomial in λ vanishes only at points where $i\lambda$ has a negative real part, that is where λ has a positive imaginary part. This completes the proof in the case where the N-dimensional $y(t)$ process has no deterministic factor. If there are such factors, it is easily verified that $G(\lambda)$ has corresponding discontinuities and the above proof then applies to $G(\lambda)$ less its jump function. The result can finally be summarized as in the statement of the theorem. If the $y(t)$ process has only deterministic factors $[R_y(t)]_{11}$ will be a sum of trigonometric functions and $G(\lambda)$ will be a function of jumps.

Conversely suppose that the $x(t)$ process has the complex spectral function (4.9.13). Then following the ideas of the proof of the analogous section of Theorem 3.9, it follows that $R(t)$ satisfies the differential equation (cf. (3.9.10) and (3.9.11)).

$$R^{(N)}(t) - a_1R^{(N-1)}(t) - \cdots - a_NR(t) = 0, \, t > 0. \tag{4.9.14}$$

Any solution of (4.9.14) is a linear combination of (at most N) functions

$$e^{\beta t}, te^{\beta t}, \cdots \tag{4.9.15}$$

where β is a root of the equation

$$\alpha^N - a_1\alpha^{N-1} - \cdots - a_N = 0 \tag{4.9.16}$$

and where powers of t may appear if β is a multiple root. Let ϵ be a positive number. The discrete parameter process determined by the variables $\{x(n\epsilon)\}$

has correlation function $R(n\epsilon)$. This function is a linear combination of function

$$(4.9.15') \qquad (e^{\beta\epsilon})^n, n(e^{\beta\epsilon})^n, \cdots$$

corresponding to those of (4.9.15). There is an equation

$$(4.9.17) \qquad \alpha^N - a_1(\epsilon)\alpha^{N-1} - \cdots - a_N(\epsilon) = 0$$

with the $\{e^{\beta\epsilon}\}$ as roots, of the same multiplicity as that of β in (4.9.16). Hence

$$(4.9.18) \qquad R[(n+N)\epsilon] - a_1(\epsilon)R[(n+N-1)\epsilon] - \cdots - a_N(\epsilon)R(n\epsilon) = 0, \quad n \geqq 0.$$

According to Theorem 3.10 the $x(n\epsilon)$ discrete parameter process is therefore a component process of an N-dimensional discrete parameter t.h.G.M. process. Since this is true for all ϵ, the $x(t)$ process is a component process of an N-dimensional continuous parameter t.h.G.M. process.

If the integral vanishes identically, the non-deterministic factors in the N-dimensional process are irrelevant to the $x(t)$ process and can be replaced by factors of type $M(0)$. If on the other hand the spectral function is continuous, the deterministic factors are irrelevant and can be replaced by factors of type M.

Proof of (*ii*). Since the t.h.G.M$_N$. processes are characterized among the component processes of N-dimensional t.h.G.M. processes by the fact that the first $N-1$ derived process exist, their spectral functions (according to Theorem 1.4) are characterized by the property that

$$\int_{-\infty}^{\infty} \lambda^{2(N-1)}\, dG(\lambda) < \infty$$

that is the numerator in (4.9.1) must be identically constant. If this constant is not 0, $\hat{G}(\lambda)$ can have no jumps, since each jump corresponds to a zero of numerator and denominator. Hence $G(\lambda)$ is either identically $\hat{G}(\lambda)$ or is in the form (4.9.2). The two possibilities obviously correspond to the deterministic and non-deterministic cases, respectively.

COROLLARY. *The t.h.G.M$_N$. one dimensional process which is the solution of* (4.4.1) *has complex spectral function*

$$(4.9.19) \qquad \int_{-\infty}^{\lambda} \frac{c^2\, d\lambda}{|(i\lambda)^N - a_1(i\lambda)^{N-1} - \cdots - a_1|^2}.$$

In fact the complex spectral function has the form (4.9.2), where the coefficients in the polynomial are those of the differential equation for the correlation function $R_y(t)$ in (4.9.9), that is the coefficients of the characteristic equation of the infinitesimal transition matrix Q, (cf. (4.4.8)). The evaluation (4.9.19) is also easily proved directly.

The analogues of Theorems 3.10 and 3.11 in the continuous parameter case are easy to prove and will be omitted.

ON CUMULATIVE SUMS OF RANDOM VARIABLES

By Abraham Wald

Columbia University

1. Introduction. Let $\{z_i\}$ (i = 1, 2, $\cdots$, ad inf.) be a sequence of independent random variables each having the same distribution. Denote by Z_j the sum of the first j elements of the sequence $\{Z_i\}$, i.e.,

$$Z_j = z_1 + z_2 + \cdots + z_j \quad (j = 1, 2, \cdots, \text{ad inf.}). \tag{1}$$

Let a be a given positive constant and b a given negative constant. Denote by n the smallest positive integer for which Z_n lies outside the open interval (b, a), i.e., Z_n is either $\leq b$ or $\geq a$. Obviously n is a random variable. If $b < Z_i < a$ for $i = 1, 2, \cdots$, ad inf., we shall say that $n = \infty$.

For any relation R we shall denote the probability that R holds by $P(R)$. It will be shown later that $P(n = \infty) = 0$, provided the variance of z_i is positive.

In this paper we shall deal with the problem of obtaining the value of $P(Z_n \geq a)$[1] and that of finding the probability distribution of n.

The study of such cumulative sums is of interest in various statistical problems. For example, a multiple sampling scheme proposed recently by Walter Bartky[2] makes use of such cumulative sums.

Cumulative sums also play an important role in the theory of the random walk of interest in physics. The results obtained in this paper may have bearing particularly on the theory of the random walk with absorbing barriers. In the presence of an absorbing wall the random walk stops whenever the particle arrives at the wall, i.e., whenever the cumulative sum of the displacements reaches a certain value.[3]

2. Two Lemmas. Lemma 1. *If the variance of z_i is not zero, $P(n = \infty) = 0$.*

Proof: Let $c = |a| + |b|$. If $n = \infty$ then for any positive integer r the following inequalities must hold

$$\left(\sum_{i=kr+1}^{(k+1)r} z_i\right)^2 < c^2 \qquad (k = 0, 1, 2, \cdots, \text{ad inf.}). \tag{2}$$

To prove $P(n = \infty) = 0$, it is sufficient to show that the probability is zero that (2) holds for all integer values of k. Since the variance of z_i is not zero, the ex-

[1] Since $P(n = \infty) = 0$, we have $P(Z_n \leqq b) = 1 - P(Z_n \geqq a)$.

[2] "Multiple sampling with constant probability", *Annals of Math. Stat.*, Vol. 14 (1943), pp. 363–377.

[3] See in this connection S. Chandrasekhar, "Stochastic problems in physics and astronomy", *Rev. of Modern Physics*, Vol. 15 (1943), p. 5.

pected value of $\left(\sum_{i=1}^{j} z_i\right)^2$ converges to ∞ as $j \to \infty$. Hence there exists a positive integer r such that

$$P\left[\left(\sum_{i=1}^{r} z_i\right)^2 < c^2\right] < 1. \tag{3}$$

From (3) it follows that the probability that (2) is fulfilled for all values of k is equal to zero. Hence $P(n = \infty) = 0$ and Lemma 1 is proved.

LEMMA 2. *Let z be a random variable such that the following four conditions are fulfilled:*

Condition I. Both the expected value Ez of z and the variance of z exist and are unequal to zero.

Condition II. There exists a positive δ such that $P(e^z < 1 - \delta) > 0$ and $P(e^z > 1 + \delta) > 0$.

Condition III. For any real value h the expected value $Ee^{hz} = g(h)$ exists.

Condition IV. The first two derivatives of the function $g(h)$ exist and may be obtained by differentiation under the integral sign, i.e.,

$$g'(h) = \frac{d}{dh} Ee^{hz} = Eze^{hz},$$

and

$$g''(h) = \frac{d^2}{dh^2} Ee^{hz} = Ez^2 e^{hz}.$$

Then there exists one and only one real value $h_0 \neq 0$ such that

$$Ee^{h_0 z} = 1.$$

PROOF: For any positive h we have

$$g(h) > P(e^z > 1 + \delta)(1 + \delta)^h. \tag{4}$$

Hence, since $P(e^z > 1 + \delta) > 0$,

$$\lim_{h=\infty} g(h) = +\infty. \tag{5}$$

Similarly we see that for any negative h

$$g(h) > P(e^z < 1 - \delta)(1 - \delta)^h. \tag{6}$$

Hence, since $P(e^z < 1 - \delta) > 0$, we have

$$\lim_{h=-\infty} g(h) = +\infty. \tag{7}$$

Since $g''(h) = Ez^2 e^{hz}$ it follows easily from Condition II that

$$g''(h) > 0, \tag{8}$$

for all real values of h.

The relations (5), (7) and (8) imply that there exists exactly one real value h^* for which $g(h)$ takes its minimum value. Since $g'(0) = Ez$ is unequal to zero by Condition I, we see that $h^* \neq 0$ and $g(h^*) < g(0) = 1$. It is clear that the function $g(h)$ is monotonically decreasing in the strict sense over the interval $(-\infty, h^*)$, and is monotonically increasing in the strict sense over the interval $(h^*, +\infty)$. Since $g(0) = 1$ and $g(h^*) < 1$, there exists exactly one real value $h_0 \neq 0$ such that $g(h_0) = 1$. Hence Lemma 2 is proved.

3. A fundamental identity. Denote by z a random variable whose distribution is equal to the common distribution of $z_i (i = 1, 2, \cdots, \text{ad inf.})$. Let D' be the subset of the complex plane such that $Ee^{zt} = \varphi(t)$ exists and is finite for any point t in D'. Consider the following identity

$$Ee^{Z_n t + (Z_N - Z_n)t} = Ee^{Z_N t} = [\varphi(t)]^N, \tag{9}$$

where N denotes a positive integer. Let P_N be the probability that $n \leq N$. For any random variable u denote by $E_N(u)$ the conditional expected value of u under the restriction that $n \leq N$, and by $E_N^*(u)$ the conditional expected value of u under the restriction that $n > N$. Then identity (9) can be written as

$$P_N E_N e^{Z_n t + (Z_N - Z_n)t} + (1 - P_N) E_N^* e^{Z_N t} = [\varphi(t)]^N. \tag{10}$$

Since in the subpopulation defined by any fixed $n \leq N$ the expression $Z_N - Z_n$ is independent of Z_n, we have

$$E_N e^{Z_n t + (Z_N - Z_n)t} = E_N e^{Z_n t} [\varphi(t)]^{N-n}. \tag{11}$$

From (10) and (11) we obtain the identity

$$P_N E_N \{e^{Z_n t} [\varphi(t)]^{N-n}\} + (1 - P_N) E_N^* e^{Z_N t} = [\varphi(t)]^N. \tag{12}$$

Dividing both sides by $[\varphi(t)]^N$ we obtain

$$P_N E_N \{e^{Z_n t} [\varphi(t)]^{-n}\} + (1 - P_N) \frac{E_N^* e^{Z_N t}}{[\varphi(t)]^N} = 1. \tag{13}$$

Let D'' be the subset of the complex plane in which $|\varphi(t)| \geq 1$ and denote by D the common part of the subsets D' and D''. Since $\lim_{N=\infty} (1 - P_N) = 0$, and since $|E_N^*(e^{Z_N t})|$ is a bounded function of N, we have in D

$$\lim_{N=\infty} (1 - P_N) \frac{E_N^* e^{Z_N t}}{[\varphi(t)]^N} = 0. \tag{14}$$

Since

$$\lim_{N=\infty} P_N E_N \{e^{Z_n t} [\varphi(t)]^{-n}\} = E\{e^{Z_n t} [\varphi(t)^{-n}\},$$

we obtain from (13) and (14) the fundamental identity

$$E\{e^{Z_n t} [\varphi(t)]^{-n}\} = 1, \tag{15}$$

for any point t in the set D.

4. Derivation of the probability that $Z_n \geq a$. In what follows in this and the subsequent sections we shall always assume that the random variable z satisfies the conditions I–IV of Lemma 2, even if this is not stated explicitly. Since it follows from Condition III that the set D' is the whole complex plane, we see that the identity (15) must hold for all points t for which $|\varphi(t)| \geq 1$.

Let $h_0 \neq 0$ be the real value for which $\varphi(h_0) = 1$. Substituting h_0 for t in (15) we obtain

$$Ee^{Z_n h_0} = 1. \tag{16}$$

Let E_1 be the conditional expected value of $e^{Z_n h_0}$ under the restriction that $Z_n \geq a$ and let E_0 be the conditional expected value of $e^{Z_n h_0}$ under the restriction that $Z_n \leq b$. Furthermore denote $P(Z_n \geq a)$ by α. Then it follows from (16)

$$\alpha E_1 + (1 - \alpha)E_0 = 1. \tag{17}$$

Hence

$$\alpha = \frac{1 - E_0}{E_1 - E_0}. \tag{18}$$

If $h_0 > 0$ then $E_1 > 1$ and $E_0 < 1$. Hence (18) implies the inequality

$$\alpha \leq \frac{1}{E_1} \leq \frac{1}{e^{ah_0}}, \qquad (h_0 > 0). \tag{19}$$

If $h_0 < 0$ then $E_1 < 1$ and $E_0 > 1$. Hence (18) implies the inequality

$$1 - \alpha \leq \frac{1}{E_0} \leq \frac{1}{e^{bh_0}}, \qquad (h_0 < 0). \tag{20}$$

We shall now derive lower and upper limits for E_0 and E_1. We derive these limits under the assumption that $h_0 > 0$. To obtain a lower limit of E_0 consider a real variable ζ which is restricted to values > 1. For any random variable u and any relation R we shall denote by $E(u \mid R)$ the conditional expected value of u under the restriction that R holds. Denote by $P(\zeta)$ the probability that $e^{h_0 z_{n-1}} < \zeta e^{bh_0}$. Then we have

$$E_0 = \int_1^\infty \left\{ \zeta e^{bh_0} E\left[e^{h_0 z} \mid e^{h_0 z} \leq \frac{1}{\zeta} \right] \right\} dP(\zeta). \tag{21}$$

Hence a lower bound of E_0 is given by

$$E_0' = e^{bh_0} \left\{ \text{g.l.b.}_{\zeta}\, \zeta E\left(e^{h_0 z} \mid e^{h_0 z} \leq \frac{1}{\zeta} \right) \right\}, \tag{22}$$

where the symbol g.l.b._{ζ} stands for greatest lower bound with respect to ζ. Since e^{bh_0} is an upper bound of E_0, we obtain the limits

$$e^{bh_0} \left\{ \text{g.l.b.}_{\zeta}\, \zeta E\left(e^{h_0 z} \mid e^{h_0 z} \leq \frac{1}{\zeta} \right) \right\} \leq E_0 \leq e^{bh_0} \qquad (h_0 > 0). \tag{23}$$

Let ρ be a real variable restricted to values > 0 and < 1. Denote by $Q(\rho)$ the probability that $e^{h_0 Z_{n-1}} < \rho e^{ah_0}$. Then similarly to (21) we obtain

$$E_1 = \int_0^1 \left\{ \rho e^{ah_0} E\left(e^{h_0 z} \mid e^{h_0 z} \geq \frac{1}{\rho} \right) \right\} dQ(\rho). \tag{24}$$

Hence an upper bound of E_1 is given by

$$e^{ah_0} \left\{ \text{l.u.b.}_{\rho}\ \rho E\left(e^{h_0 z} \mid e^{h_0 z} \geq \frac{1}{\rho} \right) \right\}. \tag{25}$$

Since e^{ah_0} is a lower bound of E_1, we obtain the following limits for E_1

$$e^{ah_0} \leq E_1 \leq e^{ah_0} \left\{ \text{l.u.b.}_{\rho}\ \rho E\left(e^{h_0 z} \mid e^{h_0 z} \geq \frac{1}{\rho} \right) \right\}, \qquad (h_0 > 0). \tag{26}$$

In a similar way upper and lower limits can be derived for E_0 and E_1 when $h_0 < 0$. With the help of these limits upper and lower limits for α can be derived on the basis of equation (18). If $h_0 > 0$ then $E_1 > 1$, $E_0 < 1$ and consequently the right hand side of (18) is a monotonically decreasing function of E_0 and E_1. Hence if E_i' is a lower, and E_i'' is an upper bound of $E_i (i = 0, 1)$, then

$$\frac{1 - E_0''}{E_1'' - E_0''} \leq \alpha \leq \frac{1 - E_0'}{E_1' - E_0'}, \qquad (h_0 > 0). \tag{27}$$

In a similar way limits for α can be obtained when $h_0 < 0$. If both the absolute value of Ez and the variance of z are small, E_0 and E_1 will be nearly equal to e^{bh_0} and e^{ah_0}, respectively. Hence, in this case a good approximation to α is given by the expression

$$\bar{\alpha} = \frac{1 - e^{bh_0}}{e^{ah_0} - e^{bh_0}}. \tag{28}$$

The difference $\bar{\alpha} - \alpha$ approaches zero if both the mean and standard deviation of z converge to zero.

5. The characteristic function of n. Let $\overline{Z}_n$ be a random variable defined as follows: $\overline{Z}_n = a$ if $Z_n \geq a$ and $\overline{Z}_n = b$ if $Z_n \leq b$. Denote the difference $\overline{Z}_n - Z_n$ by ϵ. Then ϵ is a random variable.

In what follows we shall neglect ϵ i.e., we shall substitute 0 for ϵ. No error is committed by doing so in the special case when z can take only two values d and $-d$ and the ratios a/d and b/d are integers, since in this case ϵ is exactly zero. Apart from this special case the variate ϵ will not be identical with the constant zero. However, the smaller the values $| Ez |$ and Ez^2, the smaller the error we commit by neglecting ϵ. In fact, for arbitrary small positive numbers δ_1 and δ_2 the inequality $p(| \epsilon | \leq \delta_1) \geq 1 - \delta_2$ will hold if $| Ez |$ and Ez^2 are sufficiently small. Thus in the limiting case when Ez and Ez^2 approach zero the random variable ϵ reduces to the constant zero.

(a). *The characteristic function of n when only one of the quantities a and b is finite.* It will be sufficient to treat the case when a is finite and $b = -\infty$. In this case n is defined as the smallest positive integer for which $Z_n \geq a$. To make the probability of the existence of such a value n to be equal to 1 we have to assume that the expected value μ of z is positive. Since $b = -\infty$, the fundamental identity (15) need not hold for all points t of the set D. However, it follows easily from (13) that (15) holds for all points t in D whose real part is non-negative. Denote by $\psi(\tau)$ the characteristic function of n (τ is a purely imaginary variable). Since $Z_n = a$ (neglecting ϵ), and

$$E[\varphi(t)]^{-n} = \psi[-\log \varphi(t)],$$

identity (15) can be written as

$$e^{at}\psi[-\log \varphi(t)] = 1. \tag{29}$$

Let $t(\tau)$ denote a root (with non-negative real part) of the equation in t

$$\log \varphi(t) + \tau = 0, \tag{30}$$

and substitute $t(\tau)$ for t in (29). Then we obtain

$$\psi(\tau) = e^{-at(\tau)}. \tag{31}$$

As an illustration let us calculate $\psi(\tau)$ in the case when z is normally distributed. In this case

$$\log \varphi(t) = \mu t + \frac{\sigma^2}{2} t^2,$$

where μ is the mean and σ is the standard deviation of z. Hence

$$t(\tau) = \frac{-\mu \pm \sqrt{\mu^2 - 2\sigma^2\tau}}{\sigma^2}. \tag{32}$$

If we take the + sign before the square root sign, the real part of $t(\tau)$ is non-negative, since the real part of $\sqrt{\mu^2 - 2\sigma^2\tau}$ is greater than or equal to μ. Hence the characteristic function of n is given by

$$\psi(\tau) = e^{-a/\sigma^2[-\mu+\sqrt{\mu^2-2\sigma^2\tau}]} \qquad (\mu > 0). \tag{33}$$

(b). *The characteristic function of n when a and b both are finite.* Given the value of n, let p_n be the conditional probability that $\overline{Z}_n = a$. Let p_n^* denote the probability that n is the smallest positive integer for which either $\overline{Z}_n = a$ or $\overline{Z}_n = b$ holds. Neglecting $\overline{Z}_n - Z_n$, identity (15) can be written as

$$\sum_{n=1}^{\infty} [p_n e^{at} + (1 - p_n)e^{bt}][\varphi(t)]^{-n} p_n^* = 1. \tag{34}$$

Let $\psi_1(\tau)$ be the characteristic function of n in the subpopulation where $\overline{Z}_n = a$, and let $\psi_2(\tau)$ be the characteristic function of n in the subpopulation where $\overline{Z}_n =$

b. Furthermore let $\psi(\tau)$ be the characteristic function of n in the total population.

Since we neglect the difference $\overline{Z}_n - Z_n$, it follows from (18) that the probability α that $\overline{Z}_n = a$ is given by

$$\alpha = \frac{1 - e^{bh_0}}{e^{ah_0} - e^{bh_0}}. \tag{35}$$

Putting $1 - p_n = q_n$ the following relations hold

$$\psi_1[-\log \varphi(t)] = \frac{\sum_{n=1}^{\infty} p_n p_n^* [\varphi(t)]^{-n}}{\sum p_n p_n^*} = \frac{\sum p_n p_n^* [\varphi(t)]^{-n}}{\alpha} \tag{36}$$

$$\psi_2[-\log \varphi(t)] = \frac{\sum q_n p_n^* [\varphi(t)]^{-n}}{\sum q_n p_n^*} = \frac{\sum q_n p_n^* [\varphi(t)]^{-n}}{1 - \alpha} \tag{37}$$

$$\begin{aligned} \psi[-\log \varphi(t)] &= \sum p_n^* [\varphi(t)]^{-n} = \sum (p_n + q_n)[\varphi(t)]^{-n} p_n^* \\ &= \alpha\psi_1[-\log \varphi(t)] + (1 - \alpha)\psi_2[-\log \varphi(t)]. \end{aligned} \tag{38}$$

Putting $-\log\varphi(t) = \tau$ we obtain from (34), (36) and (37)

$$\alpha\psi_1(\tau)e^{at} + (1 - \alpha)\psi_2(\tau)e^{bt} = 1. \tag{39}$$

According to Lemma 2 the equation $-\log \varphi(t) = 0$ has two different real roots in t, $t = 0$ and $t = h_0$, and $\varphi'(0)$ and $\varphi'(h_0)$ both are unequal to zero. Hence, if $\varphi(t)$ is not singular at $t = 0$ and $t = h_0$, the equation

$$-\log\varphi(t) = \tau,$$

has two roots $t_1(\tau)$ and $t_2(\tau)$ for sufficiently small values of τ such that $\lim_{\tau=0} t_1(\tau) = 0$ and $\lim_{\tau=0} t_2(\tau) = h_0$. Since the identity (15) holds for all values of t for which $|\varphi(t)| \geq 1$, and since $|\varphi[t_1(\tau)]| = |\varphi[t_2(\tau)]| = 1$ for all imaginary values of τ, it follows from (39) that both equations hold

$$\alpha\psi_1(\tau)e^{at_1(\tau)} + (1 - \alpha)\psi_2(\tau)e^{bt_1(\tau)} = 1, \tag{39'}$$

$$\alpha\psi_1(\tau)e^{at_2(\tau)} + (1 - \alpha)\psi_2(\tau)e^{bt_2(\tau)} = 1. \tag{39''}$$

Solving these two linear equations we obtain $\psi_1(\tau)$ and $\psi_2(\tau)$. The characteristic function $\psi(\tau)$ is given by

$$\psi(\tau) = \alpha\psi_1(\tau) + (1 - \alpha)\psi_2(\tau).$$

As an illustration we shall determine $\psi_1(\tau)$, $\psi_2(\tau)$ and $\psi(\tau)$ when z has a normal distribution with mean μ and standard deviation σ. We have

$$-\log \varphi(t) = -\mu t - \frac{\sigma^2}{2} t^2 = \tau.$$

Hence

$$t = \frac{-\mu \pm \sqrt{\mu^2 - 2\sigma^2\tau}}{\sigma^2}. \tag{40}$$

Putting $e^a = A$ and $e^b = B$ we obtain from (39) and (40)

$$\alpha\psi_1(\tau)A^{-\mu/\sigma^2+1/\sigma^2\sqrt{\mu^2-2\sigma^2\tau}} + (1-\alpha)\psi_2(\tau)B^{-\mu/\sigma^2+1/\sigma^2\sqrt{\mu^2-2\sigma^2\tau}} = 1 \tag{41}$$

$$\alpha\psi_1(\tau)A^{-\mu/\sigma^2-1/\sigma^2\sqrt{\mu^2-2\sigma^2\tau}} + (1-\alpha)\psi_2(\tau)B^{-\mu/\sigma^2-1/\sigma^2\sqrt{\mu^2-2\sigma^2\tau}} = 1. \tag{42}$$

These two equations are valid for any imaginary value of τ. Since $h_0 = \frac{-2\mu}{\sigma^2}$, we obtain from (35)

$$\alpha = \frac{1 - B^{-2\mu/\sigma^2}}{A^{-2\mu/\sigma^2} - B^{-2\mu/\sigma^2}}. \tag{43}$$

Let

$$g_1 = -\frac{\mu}{\sigma^2} + \frac{1}{\sigma^2}\sqrt{\mu^2 - 2\sigma^2\tau}, \tag{44}$$

and

$$g_2 = -\frac{\mu}{\sigma^2} - \frac{1}{\sigma^2}\sqrt{\mu^2 - 2\sigma^2\tau}. \tag{45}$$

Then we obtain from (41) and (42)

$$\alpha\psi_1(\tau) = \frac{B^{g_2} - B^{g_1}}{A^{g_1}B^{g_2} - A^{g_2}B^{g_1}}, \tag{46}$$

and

$$(1-\alpha)\psi_2(\tau) = \frac{A^{g_1} - A^{g_2}}{A^{g_1}B^{g_2} - A^{g_2}B^{g_1}}. \tag{47}$$

Hence the characteristic function of n is given by

$$\psi(\tau) = \frac{A^{g_1} + B^{g_2} - A^{g_2} - B^{g_1}}{A^{g_1}B^{g_2} - A^{g_2}B^{g_1}}. \tag{48}$$

6. The distribution of n when z is normally distributed. (a) *The case when a is finite and $b = -\infty$.* In this case the characteristic function of n is given by (33). Let

$$m = \frac{\mu^2}{2\sigma^2}n. \tag{49}$$

Then the characteristic function of m is given by

$$\psi^*(t) = e^{c[1-\sqrt{1-t}]}, \tag{50}$$

where

$$c = \frac{a\mu}{\sigma^2} > 0. \tag{51}$$

The distribution of m is given by

$$\frac{1}{2\pi i}\int_{-i\infty}^{i\infty} e^{c(1-\sqrt{1-t})-mt}\,dt. \tag{52}$$

Let

$$G(c, m) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} e^{-c\sqrt{1-t}-mt}\,dt, \tag{53}$$

and

$$H(c, m) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \frac{1}{\sqrt{1-t}}\, e^{-c\sqrt{1-t}-mt}\,dt. \tag{54}$$

Since

$$\frac{1}{2\pi i}\frac{d}{dt} e^{-c\sqrt{1-t}-mt} = \frac{1}{2\pi i}\left(\frac{c}{2\sqrt{1-t}} - m\right) e^{-c\sqrt{1-t}-mt} \tag{55}$$

we have

$$\frac{c}{2} H(c, m) - mG(c, m) = \frac{1}{2\pi i}[e^{-c\sqrt{1-t}-mt}]_{-i\infty}^{i\infty} = 0. \tag{56}$$

From (53) and (54) we obtain

$$\frac{\partial H(c, m)}{\partial c} + G(c, m) = 0. \tag{57}$$

From (56) and (57) it follows that

$$\frac{c}{2} H(c, m) + m\frac{\partial H(c, m)}{\partial c} = 0. \tag{58}$$

Hence

$$\log H(c, m) = -\frac{c^2}{4m} + \log \lambda(m) \tag{59}$$

where $\lambda(m)$ is some function of m only. Thus

$$H(c, m) = \lambda(m)e^{-c^2/4m}. \tag{60}$$

Now we shall determine $\lambda(m)$. We have

$$\lambda(m) = H(0, m) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \frac{1}{\sqrt{1-t}}\, e^{-mt}\,dt. \tag{61}$$

Since $(1-t)^{-1/2}$ is the characteristic function of $\frac{1}{2}\chi^2$ where χ^2 has the χ^2-distribution with one degree of freedom, the right hand side of (61) is equal to

$$\frac{1}{\Gamma(\frac{1}{2})\sqrt{m}}\, e^{-m}.$$

Hence

$$\lambda(m) = \frac{1}{\Gamma(\frac{1}{2})\sqrt{m}}\, e^{-m}. \tag{62}$$

From (60) and (61) we obtain

$$H(c, m) = \frac{1}{\Gamma(\frac{1}{2})\sqrt{m}}\, e^{-c^2/4m-m}. \tag{63}$$

From (56) and (63) we obtain

$$G(c, m) = \frac{c}{2\Gamma(\frac{1}{2})m^{3/2}}\, e^{-c^2/4m-m}. \tag{64}$$

Hence the distribution of m is given by

$$F(m)\, dm = \frac{c}{2\Gamma(\frac{1}{2})m^{3/2}}\, e^{-c^2/4m-m+c}\, dm, \qquad (0 \leq m < \infty). \tag{65}$$

Let $m = \frac{c}{2}\, m^*$. Then the distribution of m^* is given by

$$\begin{aligned} D(m^*)\, dm^* &= \frac{c^2/2}{2\Gamma\left(\frac{1}{2}\right)\left(\frac{c}{2}\right)^{3/2}(m^*)^{3/2}}\, e^{-(c/2)(1/m^*+m^*-2)}\, dm^* \\ &= \frac{\sqrt{c}}{\sqrt{2\pi}(m^*)^{3/2}}\, e^{-(c/2)(1/m^*+m^*-2)}\, dm^*. \end{aligned} \tag{66}$$

The function $\frac{1}{m^*} + m^* - 2$ is non-negative and is equal to zero only when $m^* = 1$. If c is large, then $D(m^*)$ is exceedingly small for values of m^* not close to 1. Expanding $\frac{1}{m^*} + m^* - 2$ in a Taylor series around $m^* = 1$, we obtain

$$\frac{1}{m^*} + m^* - 2 = (m^* - 1)^2 + \text{higher order terms}. \tag{67}$$

Hence for large c

$$D(m^*)\, dm^* \sim \frac{\sqrt{c}}{\sqrt{2\pi}}\, e^{-(c/2)(m^*-1)^2}\, dm^*, \tag{68}$$

i.e., if c is large m^* is nearly normally distributed with mean equal to 1 and standard deviation $\frac{1}{\sqrt{c}}$.

(b). *The case when a and b both are finite.* In this case the characteristic function of n is given by (48). Let

$$m = \frac{\mu^2}{2\sigma^2} n \qquad \text{and} \qquad d = -\frac{\mu}{\sigma^2}.$$

Then the characteristic function of m is given by

$$\psi^*(t) = \frac{A^{h_1} + B^{h_2} - A^{h_2} - B^{h_1}}{A^{h_1} B^{h_2} - A^{h_2} B^{h_1}}, \tag{69}$$

where

$$h_1 = d(1 - \sqrt{1 - t}), \qquad h_2 = d(1 + \sqrt{1 - t}), \tag{70}$$

and t is an imaginary variable. Putting $A^d = \bar{A}$, $B^d = \bar{B}$, $da = \bar{a}$ and $db = \bar{b}$, the characteristic function of m can be written as

$$\begin{aligned} \psi^*(t) &= \frac{\bar{A}(e^{-\bar{a}\sqrt{1-t}} - e^{\bar{a}\sqrt{1-t}}) + \bar{B}(e^{\bar{b}\sqrt{1-t}} - e^{-\bar{b}\sqrt{1-t}})}{\bar{A}\bar{B}(e^{(\bar{b}-\bar{a})\sqrt{1-t}} - e^{(\bar{a}-\bar{b})\sqrt{1-t}})} \\ &= \frac{\bar{A}(e^{-\bar{b}\sqrt{1-t}} - e^{(2\bar{a}-\bar{b})\sqrt{1-t}}) + \bar{B}(e^{\bar{a}\sqrt{1-t}} - e^{(\bar{a}-2\bar{b})\sqrt{1-t}})}{\bar{A}\bar{B}(1 - e^{2(\bar{a}-\bar{b})\sqrt{1-t}})}. \end{aligned} \tag{71}$$

It will be sufficient to consider only the case when $\mu > 0$, since the case < 0 can be treated in a similar way. Then $\bar{a} < 0$ and $\bar{b} > 0$. Since the real part of $+\sqrt{1 - t}$ is greater than or equal to one, we have

$$| e^{2(\bar{a}-\bar{b})\sqrt{1-t}} | < 1, \tag{72}$$

for any imaginary value of t. Let

$$T = e^{2(\bar{a}-\bar{b})\sqrt{1-t}}. \tag{73}$$

Then

$$\frac{1}{1 - T} = \sum_{j=0}^{\infty} T^j. \tag{74}$$

From (71) and (74) it follows that $\psi^*(t)$ can be written in the form of an infinite series.

$$\psi^*(t) = \sum_{i=1}^{\infty} r_i e^{-\lambda_i \sqrt{1-t}}, \tag{75}$$

where λ_i and r_i are constants and $\lambda_i > 0$. Each term of this series is a characteristic function of the form given in (50) except for a proportionality factor. Let $F_i(m)$ be the distribution of m corresponding to the characteristic function $e^{\lambda_i - \lambda_i\sqrt{1-t}}$. Then $F_i(m)$ can be obtained from (65) by substituting λ_i for c. Since we may integrate the right hand side member of (75) term by term, the distribution of m is given by

$$F(m)\, dm = \left(\sum_{i=1}^{\infty} \frac{r_i}{e^{\lambda_i}} F_i(m) \right) dm. \tag{76}$$

Since m is a discrete variable, it may seem paradoxical that we obtained a probability density function for m. However, the explanation lies in the fact that we neglected $\epsilon = \overline{Z}_n - Z_n$ and this quantity is zero only in the limiting case when μ and σ approach zero.

If $|\mu|$ and σ are sufficiently small as compared with a and $|b|$, the distribution of m given in (76) will be a good approximation to the exact distribution of m, even if z is not normally distributed. The reason for this can be indicated as follows: Let

$$z_i^* = \sum_{j=(i-1)r+1}^{ir} z_j \qquad (i = 1, 2, \cdots, \text{ad inf.}) \tag{77}$$

where r is a given positive integer. Since the variates z_j are independently distributed each having the same distribution, under some weak conditions the variates $z_i^* (i = 1, 2, \cdots, \text{ad inf.})$ will be nearly normally distributed for large r. Hence, considering the cumulative sums $Z_i^* = z_1^* + z_2^* + \cdots + z_i^* (i = 1, 2, \cdots, \text{ad inf.})$, the distribution given in (76) is applicable with good approximation, provided that $r|\mu|$ and $\sqrt{r}\sigma$ are small as compared with a and $|b|$ so that the difference $\epsilon^* = \overline{Z}_n^* - Z_n^*$ can be neglected.

7. The exact probability distribution of Z_n and the exact characteristic function of n when z can take only integral multiples of a given constant d. In the previous sections we derived the probability $P(Z_n \geq a)$ and the characteristic function of n under the assumption that the quantity by which Z_n may differ from a or b is small and can be neglected. This can be done whenever $|Ez|$ and Ez^2 are small. However, if $|Ez|$ or Ez^2 is not small, it is desirable to derive the exact probability distribution of Z_n and the exact characteristic function of n. Both are obtained in the present section for random variables z which can take only a finite number of integral multiples of a given constant d. This is a rather general result, since any distribution of z can be approximated arbitrarily fine by a discrete distribution of the above type if the constant d is chosen sufficiently small.

There is no loss of generality in assuming that $d = 1$, since the quantity d can be chosen as the unit of measurement. Thus, we shall assume that z takes only a finite number of integral values. Let g_1 and g_2 be two positive integers such that $P(z = -g_1)$ and $P(z = g_2)$ are positive and z can take only integral values $\geq -g_1$ and $\leq g_2$. Denote $P(z = i)$ by h_i. Then the characteristic function of z is given by

$$\varphi(t) = \sum_{i=-g_1}^{g_2} h_i e^{ti}. \tag{78}$$

To obtain the roots of the equation $\varphi(t) = 1$, we put $e^t = u$ and solve the equation

$$\sum_{i=-g_1}^{g_2} h_i u^i = 1. \tag{79}$$

Denote $g_1 + g_2$ by g and let the g roots of (79) be $u_1, \cdots, u_g$, respectively. We shall assume that no two roots are equal, i.e., $u_i \neq u_j$ for $i \neq j$. Substituting u_i for e^t in the identity (15) we obtain

$$E(u_i^{Z_n}) = 1 \qquad (i = 1, \cdots, g). \tag{80}$$

Denote by $[a]$ the smallest integer $\geq a$, and by $[b]$ the largest integer $\leq b$. Then Z_n can take only the values

$$[b] - g_1 + 1, \quad [b] - g_1 + 2, \cdots, [b], [a], \quad [a] + 1, \cdots, [a] + g_2 - 1. \tag{81}$$

Denote the g different integers in (81) by $c_1, \cdots, c_g$, respectively. Furthermore, denote $P(Z_n = c_i)$ by ξ_i. Then equations (80) can be written as

$$\sum_{j=1}^{g} \xi_j u_i^{c_j} = 1 \qquad (i = 1, \cdots, g). \tag{82}$$

Let Δ be the determinant value of the matrix $\| u_i^{c_j} \|$ $(i, j = 1, \cdots, g)$ and let Δ_j be the determinant we obtain from Δ by substituting 1 for the elements in the jth column. If $\Delta \neq 0$, it follows from (82) that $P(Z_n = c_j) = \xi_j$ is given by

$$\xi_j = \frac{\Delta_j}{\Delta}. \tag{83}$$

Hence, $P(Z_n \geq a) = \sum_j (\Delta_j/\Delta)$ summed for all values of j for which $c_j \geq a$. From the probability distribution of Z_n we can easily derive the expected value En of n. In fact, differentiating the fundamental identity (15) with respect to t at $t = 0$ we obtain

$$E\left[Z_n - \frac{\varphi'(0)}{\varphi(0)}\, n\right] = 0. \tag{84}$$

Since $\dfrac{\varphi'(0)}{\varphi(0)} = Ez$, we obtain from (84)

$$En = \frac{EZ_n}{Ez} = \frac{1}{Ez} \sum_{j=1}^{g} \frac{c_j \Delta_j}{\Delta}. \tag{85}$$

Now we shall derive the exact characteristic function of n. Denote by $\psi_i(\tau)$ (τ is a purely imaginary variable) the characteristic function of the conditional distribution of n under the restriction that $Z_n = c_i$. Let $t_1(\tau), \cdots, t_g(\tau)$ be g roots of the equation

$$\varphi(t) = e^{-\tau}, \tag{86}$$

such that

$$\lim_{\tau=0} e^{t_i(\tau)} = u_i. \tag{87}$$

Substituting $t_i(\tau)$ for t in the fundamental identity (15) we obtain

$$\sum_{j=1}^{g} \xi_j e^{c_j t_i(\tau)} \psi_j(\tau) = 1 \qquad (i = 1, \cdots, g). \tag{88}$$

These equations are linear in the unknowns $\psi_1(\tau), \cdots, \psi_g(\tau)$ and the determinant of these equations is given by

$$\delta(\tau) = \begin{vmatrix} \xi_1 e^{c_1 t_1(\tau)} & \cdots & \xi_g e^{c_g t_1(\tau)} \\ \xi_1 e^{c_1 t_2(\tau)} & \cdots & \xi_g e^{c_g t_2(\tau)} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \xi_1 e^{c_1 t_g(\tau)} & \cdots & \xi_g e^{c_g t_g(\tau)} \end{vmatrix}. \tag{89}$$

Obviously, $\delta(0) = \xi_1 \xi_2 \cdots \xi_g \Delta$. Hence if $\xi_i \neq 0$ $(i = 1, \cdots, g)$ and $\Delta \neq 0$, also $\delta(0) \neq 0$ and consequently $\delta(\tau) \neq 0$ for any τ with sufficiently small absolute value. Thus, $\psi_1(\tau), \cdots, \psi_g(\tau)$ can be obtained by solving the linear equations (88). The characteristic function $\psi(\tau)$ of the unconditional distribution of n is given by

$$\psi(\tau) = \sum_{i=1}^{g} \xi_i \psi_i(\tau). \tag{90}$$

SOME IMPROVEMENTS IN WEIGHING AND OTHER EXPERIMENTAL TECHNIQUES[1]

By Harold Hotelling

Columbia University

When several quantities are to be ascertained there is frequently an opportunity to increase the accuracy and reduce the cost by combining suitably in one experiment what might ordinarily be considered separate operations. The theory of design of experiments developed as a branch of modern mathematical statistics, and of which fundamental considerations are set forth in R. A. Fisher's book [1], provides many improvements of this kind. Since the main interests of Fisher and other originators of this theory have been in biology, the applications so far made have been chiefly biological in character, excepting for certain economic and social investigations involving stratified sampling. The possibilities of improvement of physical and chemical investigations through designed experiments based on the theory of statistical inference have scarcely begun to be explored.

The following example is due to F. Yates [2]. A chemist has seven light objects to weigh, and the scale also requires a zero correction, so that eight weighings are necessary. The standard error of each weighing is denoted by σ, the variance therefore by σ^2. Since the weight assigned to each object by customary techniques is the difference between the reading of the scale when carrying that object and when empty, the variance of the assigned weight is $2\sigma^2$, and its standard error is $\sigma\sqrt{2}$.

The improved technique suggested by Yates consists of weighing all seven objects together, and also weighing them in groups of three so chosen that each object is weighed four times altogether, twice with any other object and twice without it. Calling the readings from the scale $y_1, \cdots, y_8$ we then have as equations for determining the unknown weights $a, b, \cdots, g$,

$$\begin{aligned}
a + b + c + d + e + f + g &= y_1 \\
a + b + c &= y_2 \\
a + d + e &= y_3 \\
a + f + g &= y_4 \\
b + d + f &= y_5 \\
b + e + g &= y_6 \\
c + d + g &= y_7 \\
c + e + f &= y_8 .
\end{aligned}$$

[1] Presented at the Wellesley meeting of the Institute of Mathematical Statistics, Aug. 13, 1944.

Any particular weight is found by adding together the four equations containing it, subtracting the other four, and dividing by 4. Thus

$$a = \frac{y_1 + y_2 + y_3 + y_4 - y_5 - y_6 - y_7 - y_8}{4}.$$

The variance of a sum of independent observations is the sum of their variances, as is well known, and the variance of c times an observation is c^2 times the variance of that observation. Taking $c = \frac{1}{4}$ for the first four terms in the expression for a and $c = -\frac{1}{4}$ for the others gives for the variance of a by this method $\sigma^2/2$, which is only one-fourth that for the direct method. The standard error, or probable error, has been halved. If a degree of accuracy is required calling for repetition a certain number of times of the weighings by the direct method, then only one-fourth as many weighings are needed by Yates' method to procure the same accuracy in the average.

A further improvement, which does not seem to have been mentioned in the literature, will be obtained if Yates' procedure is modified by placing in the other pan of the scale those of the objects not included in one of his weighings. Calling the readings in this case $z_1, \cdots, z_8$, we have

$$\begin{aligned}
a + b + c + d + e + f + g &= z_1 \\
a + b + c - d - e - f - g &= z_2 \\
a - b - c + d + e - f - g &= z_3 \\
a - b - c - d - e + f + g &= z_4 \\
-a + b - c + d - e + f - g &= z_5 \\
-a + b - c - d + e - f + g &= z_6 \\
-a - b + c + d - e - f + g &= z_7 \\
-a - b + c - d + e + f - g &= z_8 .
\end{aligned}$$

From these equations,

$$a = \frac{z_1 + z_2 + z_3 + z_4 - z_5 - z_6 - z_7 - z_8}{8},$$

with a like expression for each of the other unknowns. The variance of each unknown by this method is $\sigma^2/8$. The standard error is half that by Yates' method, or a quarter of its value by the direct method of weighing each object separately. The number of repetitions required to procure a particular standard error in the mean is one-sixteenth that by the direct method.

A simpler example illustrating the same point is that of two objects to be weighed, with a scale already corrected for bias. Again let σ^2 be the variance of an individual weighing. If we weigh the two objects together in one pan of the scale, and then in opposite pans, we have as equations for the unknown weights a and b,

$$a + b = z_1, \qquad a - b = z_2,$$

whence

$$a = (z_1 + z_2)/2, \qquad b = (z_1 - z_2)/2.$$

The variances of a and b by this method are both equal to $\sigma^2/2$, half the value when the two objects are weighed separately. The means found from a number of pairs of weighings of sums and differences have the same precision as those found from twice as many pairs of weighings of the objects separately.

Further economies of effort, or gains in accuracy, are possible with larger numbers of weighings and of objects to be weighed. These improvements can to some extent be applied also to other types of measurement, as of distances, since it is sometimes possible to measure the sum of a number of such quantities, or the difference between two such sums, with approximately the same accuracy as a single one of them. The outstanding case, however, seems to be that of weighing on a balance objects light enough so that their aggregate weight is below the maximum for which the balance was designed, since in this case it is quite reasonable to assume that the several recorded results all have the same standard error σ and that they are independent.

In what follows, some principles underlying the design of efficient schemes of this kind will be developed and applied to obtain some additional plans. However no comprehensive general solution has been reached; this appears to be a matter for further mathematical research. Also, we leave aside in this paper the problem of estimating the error variance. All this discussion is based on the minimization of the actual variance. In order to utilize the results it is necessary that this variance be either known a priori or estimated from the residuals from the least-square solution. The latter type of estimate is in some ways more satisfactory, since it refers to the actual experiment rather than to some previous experiments which may not have been made under exactly the same conditions. But in order to have such an estimate it is necessary that the number of observations exceed the number of unknowns, and desirable that the excess shall have a large enough value to insure a stable estimate of the error variance σ^2. The appropriate test for significance, or determination of confidence limits for the unknowns, must then utilize the Student distribution or its generalization, the variance ratio distribution, which take full account of the instability caused by an inadequate number of degrees of freedom for estimating σ.

It is only when σ is known exactly apart from the experiment being designed that the criteria we here consider are exactly applicable. In other cases there may need to be a balancing, in the design of the experiment, between the desiderata of *minimum* variance and of *accurately known* variance, with the accuracy of this knowledge depending on the number of available degrees of freedom. A theory of design taking full account of this consideration would require a use of the power functions of the Student distribution and the variance ratio distribution, discovered respectively by R. A. Fisher [3] and P. C. Tang [4].

We shall denote by N the number of weighings to be made, and by p the number of objects to be weighed. In order that it be possible to determine the un-

known weights from the observations it is necessary that $p \leqq N$, and if a possible bias in the scale must be eliminated by means of the same data it is necessary that $p \leqq N - 1$. Supposing these conditions to be satisfied, we shall show, among other things, that the minimum possible variance for one of the unknowns is σ^2/N; that the experiment may be arranged so that a selected one of the unknowns has exactly this minimum variance excepting when N is odd and a bias must be allowed for also; and that for some, but not all, combinations of p and N, this minimum variance is attained for all the unknowns simultaneously. This minimum value σ^2/N is of course equal to the variance of the mean of N weighings of one object alone, disregarding the rest; but it will be seen below that by complex experiments of the kind indicated, determinations from the same number of weighings of the other weights also can at the same time be made with some finite variance, which may or may not have the minimum value.

The following notation will be used in the proof. Let $x_{i\alpha} = 1$ or -1 if the ith object is included in the αth weighing by being placed respectively in the left- or right-hand pan, and let $x_{i\alpha} = 0$ if the ith object is not included in the αth weighing. Here $i = 1, 2, \cdots, p$ and $\alpha = 1, \cdots, N$. Let y_α be the result recorded for the αth weighing, let Δ_α be the error in this result, and let b_i be the true weight of the ith object, so that we have the N equations

$$x_{1\alpha}b_1 + x_{2\alpha}b_2 + \cdots + x_{p\alpha}b_p = y_\alpha + \Delta_\alpha , \tag{1}$$

provided there is no bias, or if by y_α we mean the observed weight corrected for a bias known a priori. Under these conditions the estimate of each of the b_i's having the properties of zero bias and minimum variance is that provided by the method of least squares. This statement, which does not depend on any assumption of a normal or other particular form of distribution of the errors, has been known long but not widely, since there is an easier derivation of the method by the application to the normal distribution to the method of maximum likelihood. Its proof, due originally to Laplace, has appeared in many forms in the work of Gauss and later authors [5]; the latest version is by the present writer [6].

Letting S stand for summation over all the N weighings we put

$$a_{ij} = Sx_{i\alpha}x_{j\alpha} , \qquad g_i = Sx_{i\alpha}y_\alpha , \tag{2}$$

and write the normal equations in the form

$$\Sigma a_{ij}b_j = g_i ,$$

where Σ stands for a sum with respect to j from 1 to p. From the usual theory of least squares (cf. for example the reference last cited) it is known that the standard error of the determination of b_1 from these equations—which is the

minimum possible standard error of b_1 for any way of combining the observations—is σ times the square root of A_{11}/A, where

$$A = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \cdot & \cdot & \cdot & \cdot \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{vmatrix},$$

and A_{11} is the minor of A obtained by deleting the first row and column.

The matrices of A and of A_{11} are known to be positive definite or semi-definite. The semi-definite case is excluded by the consideration that the normal equations shall actually determine the unknowns. Hence the inverse of the latter matrix exists and is positive definite. But this inverse, which we may write

$$d = \begin{bmatrix} d_{22} & \cdots & d_{2p} \\ \cdot & \cdot & \cdot \\ d_{p2} & \cdots & d_{pp} \end{bmatrix},$$

consists of the coefficients in the identity

$$A/A_{11} = a_{11} - \sum_{i,j=2}^{p} d_{ij} a_{i1} a_{j1}.$$

which is obtained by expanding A with reference to its first row and first column. The positive definite character of d therefore leads to the following

LEMMA: *If $a_{12}, \cdots, a_{1p}$ $(= a_{21}, \cdots, a_{p1}$ respectively) are free to vary while the other elements of A remain fixed, the maximum value of A/A_{11} is a_{11}, and is attained when and only when $a_{12} = a_{13} = \cdots = a_{1p} = 0$.*

From this it is evident that the variance of b_1, namely $\sigma^2 A_{11}/A$, cannot be less than σ^2/a_{11}, and will reach this value only if the experiment is so arranged that the elements after the first in the first row and column of A are all zero. That such an arrangement is possible may be seen by a consideration of the matrix

$$X = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{p1} \\ x_{12} & \cdot & \cdot & x_{p2} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x_{1N} & \cdot & \cdot & x_{pN} \end{bmatrix}$$

whose elements are restricted to be 1's, -1's and 0's. The condition $a_{i1} = 0$, by (2), means simply that $Sx_{i\alpha}x_{1\alpha} = 0$, a condition which may be expressed by saying that the first column of X is orthogonal to the ith column. The condition that the variance of b_1 have its minimum value σ^2/a_{11} is thus, according to the lemma, that the first column of X shall be orthogonal to all the others. The *minimum minimorum* of this variance will be reached if the first row of X is

not only orthogonal to all the others, but consists entirely of 1's and -1's, so that $a_{11} = N$. The value of this minimum minimorum is σ^2/N.

If there is a possible bias b_0 this procedure needs to be modified by the addition of b_0 to the left member of (1) and subsequent treatment of this term like the others, putting $x_{0\alpha} = 1$ in (2), and modifying X by adjoining a column of 1's. The necessary and sufficient condition that the variance of b_1 shall equal σ^2/N is then that the column

$$\begin{matrix} x_{11} \\ x_{12} \\ \cdots \\ x_{1N} \end{matrix}$$

shall consist entirely of 1's and -1's and shall be orthogonal to a column consisting of 1's, and to all the other columns of X.

If no bias needs to be eliminated the experiment can be arranged so that the variance of b_1 is σ^2/N merely by filling up the first column of X with 1's and -1's in any arbitrary manner, and then choosing the later columns so as to be orthogonal to this first one, and so that all are linearly independent. This can be accomplished, for example by choosing the first element in all the columns to be the same as that in the first column; choosing the ith element in the ith column ($i = 2, 3, \cdots, p$) to be the negative of the ith element in the first column; and making all the other elements of X zero.

When a bias is to be eliminated, so that there is a column of 1's in X corresponding to b_0, it is necessary that N be even in order that the column of X corresponding to b_1 may consist of 1's and -1's in equal numbers, without any 0's, a condition essential for the orthogonality between these two columns with the maximum value N for a_{11}. Supposing N even, let us assign the value 1 to each of the first $N/2$ elements of the column corresponding to b_1 and the value -1 to the last $N/2$ elements of this row. The remaining rows of X may then be filled up by the same method as that indicated above for the case in which there is no bias. The variance of b_1 will then take its theoretical minimum value σ^2/N.

If N is odd and there is a possible bias, the column of X corresponding to b_1 can be filled up with 1's and -1's in equal numbers, with a single zero, and the remaining columns can be made orthogonal to it. The variance of b_1 in this case will be $\sigma^2/(N - 1)$.

The method suggested above for filling up the later columns of X is convenient for the proof, but is not usually to be recommended in practice, since other methods will in all but the simplest cases give smaller standard errors for the unknowns other than the first. For some values of N and p it is possible to determine all the unknowns with equal and minimum variance. These are the cases in which all the columns of X can be made mutually orthogonal and

without zeros, excepting that the column corresponding to b_0 may contain some zeros. Thus for $N = 4$ the scheme of weighing represented by the matrix

$$X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix},$$

whose columns are all mutually orthogonal, may be applied to weigh three objects when there is a possible bias, or four where there is not, with variance $\sigma^2/4$ for each of the unknowns in either case. The matrix $X'X$ of the normal equations has the form

$$\begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}.$$

Calling the results of the weighings y_0, y_1, y_2, y_3 in the case of possible bias we have for the unknowns the expressions

$$b_1 = (y_0 + y_1 - y_2 - y_3)/4$$

$$b_2 = (y_0 - y_1 + y_2 - y_3)/4$$

$$b_3 = (y_0 - y_1 - y_2 + y_3)/4.$$

The complete orthogonality exemplified by this design has several advantages besides the fact that the variance of each of the unknowns has the same minimum value as if all the weighings were to be devoted to it alone (or half the value of the variance of this unknown if half the weighings were devoted to it plus bias and half to determining the bias). The diagonal form of the matrix $X'X$ means that the labor of solving normal equations, which is sometimes formidable, is reduced to the trivial task of dividing by N. Also, the diagonal form of this matrix implies that its inverse is also of diagonal form, from which it follows that the estimates of the different unknowns are statistically independent. Consequently the variances, or standard errors, of linear functions of the unknowns are easy to find. Thus the variance of the difference between the estimates of two of the weights is simply the sum of their variances. But of course if the main object of the experiment is to determine a particular difference of this kind, or any other linear function of the weights, a different design should be sought to minimize the particular variance which is of interest.

In contrast to the satisfactory design possible with four weighings, no complete orthogonality is possible with six weighings, or with any odd number, if the number of objects to be weighed is the maximum possible for the number of

weighings and if each object is actually to enter into each weighing in one pan or the other. For $N = 3$ and bias known to be zero consider the scheme

$$X = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{bmatrix},$$

which corresponds to weighing two objects, first together in one pan, then in opposite pans, and then weighing one alone. Calling b_1 the weight of the object that has been on the scale through all three weighings and b_2 the other we have the estimates

$$b_1 = (y_1 + y_2 + y_3)/3$$
$$b_2 = (y_1 - y_2)/2,$$

with respective variances

$$\sigma_1^2 = \sigma^2/3, \qquad \sigma_2^2 = \sigma^2/2.$$

Thus the first weight is determined with the minimum possible variance but the second is not.

An alternative method of weighing under these same conditions is to weigh both objects in one pan together twice and to weigh them in opposite pans once. This gives

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix},$$

with the normal equations

$$3b_1 + b_2 = y_1 + y_2 + y_3$$
$$b_1 + 3b_2 = y_1 + y_2 - y_3,$$

whose solution is

$$b_1 = (y_1 + y_2 + 2y_3)/4$$
$$b_2 = (y_1 + y_2 - 2y_3)/4,$$

and variances

$$\sigma_1^2 = \sigma_2^2 = \tfrac{3}{8}\sigma^2.$$

Thus the weights are by this method determined with equal accuracy, which is better than by the preceding method for one of the objects but worse for the other. To choose between the two methods it is therefore appropriate to take into consideration the relative accuracy desired in the weights of the two objects. Either method is better than weighing the objects separately.

Either of these two X matrices can also be made the basis for weighing a single

object when the scale is suspected of having a bias. The weight of this object will be estimated as b_2, and will have the variance $\frac{1}{2}\sigma^2$ by the first method, or $\frac{3}{5}\sigma^2$ by the second. Thus the second method is distinctly superior in this case.

Orthogonality between columns obviously requires both negative and positive signs, corresponding to weighings in both pans of the balance. Thus the experimental designs of maximum efficiency for weighing on a balance are not available with a spring scale, or in making measurements of any kind in which it is not possible to arrange that the quantities read off are differences. In such cases the elements of X are restricted to be 1 or 0. Let us now consider some of the simplest cases of this kind, assuming for simplicity that $\sigma = 1$. We shall deal only with cases in which there is no bias.

For $N = 3$, $p = 2$ the simple experiment of weighing one object twice and the other once yields variances $\frac{1}{2}$ and 1 respectively. All other designs are in this case less satisfactory, with the possible exception of that specified by

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix},$$

with $b_1 = (y_1 + 2y_2 - y_3)/3$ and $b_2 = (y_1 - y_2 + 2y_3)/3$ having each the variance of $\frac{2}{3}$.

For $N = 3$, $p = 3$ the most efficient design is given by

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix},$$

with $b_1 = (y_1 + y_2 - y_3)/2$, and b_2 and b_3 given by cyclic permutation in this formula. The variance of each unknown is $\frac{3}{4}$.

For $N = 4$, $p = 3$ a design having an advantage in some situations is that given by

$$X = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

(together of course with those obtained by permutations of rows and of columns, as is to be understood throughout). The normal equations are

$$\begin{aligned} 3b_1 + 2b_2 + 2b_3 &= y_1 + y_2 + y_3 \\ 2b_1 + 3b_2 + 2b_3 &= y_2 + y_3 + y_4 \\ 2b_1 + 2b_2 + 3b_3 &= y_1 + y_3 + y_4 . \end{aligned}$$

An expeditious method of solution in this as in many similar cases is to add them all together and then subtract an appropriate multiple of the sum from each of the normal equations in turn. The variance of each unknown found by this

experiment is $5/7 = .714$. The simple experiment consisting of weighing one of the objects twice and the others once each yields variances in one case larger and in two cases smaller than this.

For $N = p = 4$ the cyclic arrangement

$$\begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

leads to variances all equal to $7/9$.

For $N = 5, p = 2$ the most efficient design appears to be

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

The variance of each unknown is in this case $1/3$.

REFERENCES

[1] R. A. Fisher, *The Design of Experiments*, Oliver and Boyd, London and Edinburgh, 1935 and 1937.

[2] F. Yates, "Complex experiments," *Jour. Roy. Stat. Soc. Supp.*, Vol. 2 (1935), pp. 181-247, including discussion. The reference is to p. 211.

[3] R. A. Fisher, "The sampling error of estimated deviates. . . ," *Mathematical Tables*, Vol. 1 (1931), pp. xxvi–xxxv. London, British Association for the Advancement of science.

[4] P. C. Tang, "The power function of the analysis of variance tests with tables and illustrations of their use," *Stat. Res. Mem.*, Vol. 2 (1939), pp. 126–149.

[5] Cf. Whittaker and Robinson, *Calculus of Observations*, Section 115.

[6] Harold Hotelling, "Problems of prediction," *Am. Jour. of Sociology*, Vol. 48 (1942), pp. 61–76.

ON THE ANALYSIS OF A CERTAIN SIX-BY-SIX FOUR-GROUP LATTICE DESIGN[1]

By Boyd Harshbarger[2]

Virginia Agricultural Experiment Station

1. Introduction. The lattice consists of groups of randomized incomplete blocks with certain restrictions being imposed on the randomization within each group, and the number of varieties is a perfect square. For example, if the number of varieties is $k^2 = 36$, then the orthogonal groups for a triple lattice, not considering randomizing within the blocks or between blocks, are as follows: (the numbers signify varieties).

GROUP X

Blocks						
(1)	1	2	3	4	5	6
(2)	7	8	9	10	11	12
(3)	13	14	15	16	17	18
(4)	19	20	21	22	23	24
(5)	25	26	27	28	29	30
(6)	31	32	33	34	35	36

GROUP Y

Blocks						
(1)	1	7	13	19	25	31
(2)	2	8	14	20	26	32
(3)	3	9	15	21	27	33
(4)	4	10	16	22	28	34
(5)	5	11	17	23	29	35
(6)	6	12	18	24	30	36

GROUP Z

Blocks						
(1)	1	8	15	22	29	36
(2)	2	9	16	23	30	31
(3)	3	10	17	24	25	32
(4)	4	11	18	19	26	33
(5)	5	12	13	20	27	34
(6)	6	7	14	21	28	35

This design is constructed so that no variety appears with another variety more than once in the same block. This important characteristic makes the analysis simple, as it enables the results to be treated as a factorial design. The analysis is well described by Yates [3, 4, 5] and Cochran [1].

Suppose another group, U, is now formed from a six by six lattice, for example, the following group:

---

[1] Certain of the ideas presented here are embodied in the author's unpublished doctoral thesis by the same title, Library, George Washington University, Washington, D. C., 1943.

[2] The author wishes to express his appreciation to W. G. Cochran of Iowa State College, who advised freely in the preparation of the original thesis, and to Frank M. Weida of George Washington University.

GROUP U

Blocks						
(1)	31	26	21	16	11	6
(2)	1	32	27	22	17	12
(3)	7	2	33	28	23	18
(4)	13	8	3	34	29	24
(5)	19	14	9	4	35	30
(6)	25	20	15	10	5	36

The important characteristic, that no variety appears with another variety in the same block more than once, does not hold. For example, varieties 1 and 22 appear together in both groups Z and U.

It is the purpose of this paper to develop the statistical method for the analysis of such a design, where each group is duplicated, and to apply the results to an actual problem. The least square solution, as developed here, uses only the intra-block information to correct the varieties for block effects. In a second article the solution using both intra- and inter-block information will be given.

2. Estimation of the block and varietal effects. It is reasonable to assume in varietal trials that the general mean, and any particular block and variety effects, operate additively to produce the true mean of y associated with this block and variety. In particular, if y_{eij} is the yield of the plot for the jth variety in the ith block of the eth replicate, the following hypothesis may be set up, namely:

$$y_{eij} = \mu + \rho_e + \beta_{ei} + \nu_j + \epsilon_{eij}. \tag{1}$$

Where μ is the true or population mean yield, ρ_e is the population replicate effect of the eth replicate, β_{ei} is the population block effect of the ith block in the eth replicate, ν_j is the population variety effect of the jth variety, and ϵ_{eij} is the experimental error of the eij plot. Since the design has eight replicates, that is, each group is duplicated, the block effects may be estimated from unpaired and paired replicates or partners.

It is assumed that the ϵ_{eij} are independently and normally distributed with common variance $\frac{1}{W}$. Without loss of generality, it also may be assumed that the sum of the replicate effects, the sum of the block effects within any replication, and the sum of the variety effects are respectively equal to zero.

The parameters are estimated by the method of least squares, subject to the restrictions stated in the preceding paragraph. This involves choosing the parameters so that

$$S\left(y_{eij} - mx_1 - r_e x_{2e} - \frac{b_{ei} - b'_{ei}}{2} x_{3ei} - \frac{b_{ei} + b'_{ei}}{2} x_{4ei} - v_j x_{5j}\right)^2 + \lambda_1 \sum_{1}^{8} r_e + \sum_{e=1}^{8} \lambda_{2e} \sum_{i=1}^{6} \frac{b_{ei} - b'_{ei}}{2} + \sum_{e=1}^{8} \lambda_{3e} \sum_{i=1}^{6} \frac{b_{ei} + b'_{ei}}{2} + \lambda_4 \sum_{j=1}^{36} V_j \tag{2}$$

is a minimum.[3] Here y_{eij} is the dependent variate, and $x_1, \cdots, x_{5j}$ are the independent variates. In ordinary regression problems, the values of the x variates, as well as the y variate, constitute a part of the original data. However, in this case the y variate only is given, and the x variates must be constructed. Thus, for the design, one takes $x_1 = 1$ for all values; $x_{2e} = 1$ for all values in replicate e, but zero elsewhere; $x_{3ei} = 1$ for all values in the ith block of the eth replicate and -1 for all values in its partner, but zero elsewhere; $x_{4ei} = 1$ for all values in the ith block of the eth replicate and also 1 in its partner, but zero elsewhere; and $x_{5j} = 1$ where variety j occurs, but zero elsewhere.

One now takes the partial derivatives of the above equation with respect to the parameters and forms the normal equations. It can be shown that $(\lambda_1, \cdots, \lambda_4)$ are each zero. The normal equations not involving λ's are:

Leading term

m $$Nm + k^2 \sum_{e=1}^{8} r_e + 2k \sum_{e=1}^{8} \sum_{i=1}^{6} \frac{b_{ei} - b'_{ei}}{2} + 2k \sum_{e=1}^{8} \sum_{i=1}^{6} \frac{b_{ei} + b'_{ei}}{2} + 8 \sum_{j=1}^{36} V_j = G.$$

(3) r_e $$k_2 m + k^2 r_e + k \sum_{i=1}^{6} b_{ei} + \sum_{j=1}^{36} v_j = R_e.$$

$\frac{b_{ei} - b'_{ei}}{2}$ $$k(r_e - r'_e) + 2k \frac{(b_{ei} - b'_{ei})}{2} = B_{ei} - B'_{ei}.$$

Equations having $\frac{b_{ei} + b'_{ei}}{2}$ as leading terms.

Equations having v_j as leading terms.

In the above, N is the total number of values, k is the number of plots in a block, r_e is the eth replicate effect, b_{ei} is the ith block effect in the eth replicate, v_j is the jth variety effect, G is the total sum of all values, R_e is the sum of the values in the eth replicate, B_{ei} is the sum of values in the ith block of the eth replicate and v_j is the sum of the yields of the jth variety. The primes denote similar values of the partner terms.

Using the restrictions

$$\text{(4)} \qquad \sum_{e=1}^{8} r_e = \sum_{e=1}^{8} \sum_{i=1}^{6} \frac{b_{ei} + b'_{ei}}{2} = \sum_{e=1}^{8} \sum_{i=1}^{6} \frac{b_{ei} - b'_{ei}}{2} = \sum_{j=1}^{36} v_j = 0,$$

the values of the following parameters are directly obtainable:

$$\text{(5)} \quad m = \frac{G}{N}, \qquad r_e = \frac{R_e}{k^2} - \frac{G}{N} \quad \text{and} \quad \frac{b_{ei} - b'_{ei}}{2} = \frac{B_{ei} - B'_{ei}}{2k} - \frac{r_e - r'_e}{2}.$$

[3] S, for summation, will be used to represent summation over all values. Σ will be used in a more restricted sense.

The values of the $\frac{b_{ei} + b'_{ei}}{2}$ and v_i effects cannot be obtained directly. In order to simplify the solution of these equations, only the mean, confounded blocks, and the varietal effects will be used. The results later will be corrected for replicate effects. If each of the yields is added to the corresponding yield in its partner, one gets an equation of the form

$$y_{ik} = \mu + \beta_i + \nu_k + \epsilon_{ik} \tag{6}$$

where μ, β_i, and ν_k are now twice their original values. These parameters are estimated by the method of least squares subject to the restrictions previously given. To distinguish them from the estimates derived in equation (3) they are designated with double primes ($''$). If B_{ix} is the total for block i in group X, that is from both pairs, and similarly for the other groups, the normal equations are:

Leading term

$$\begin{array}{lll}
m'' & 144m'' & = G. \\
b''_{x1} & 6m'' + 6b''_{x1} + (v''_1 + v''_2 + v''_3 + v''_4 + v''_5 + v''_6) & = B_{1x}. \\
 & \cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\
b''_{u6} & 6m'' + 6b''_{u6} + (v''_5 + v''_{10} + v''_{15} + v''_{20} + v''_{25} + v''_{36}) & = B_{u6}. \\
v''_1 & 4m'' + b''_{x1} + b''_{y1} + b''_{z1} + b''_{u2} + 4v''_1 & = V_1. \\
 & \cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\
v''_{36} & 4m'' + b''_{x6} + b''_{y6} + b''_{z1} + b''_{u3} + 4v''_{36} & = V_{36}.
\end{array} \tag{7}$$

Let T_{xi} be the total of all the varieties appearing in block 1 of group X from all the replicates; T_{yi} the total of all values appearing in block i of group Y from all the replicates, etc. Also

$$C_{xi} = 4B_{xi} - T_{xj} \qquad C_{yi} = 4B_{yi} - T_{yj} \quad \text{etc.} \tag{8}$$

Solving the equations:

$$b''_{xi} = \frac{C_{xi}}{18} \qquad b''_{yi} = \frac{C_{yi}}{18}$$

and

$$\begin{array}{ll}
b''_{z1} = \frac{1}{432}[(25C_{x1} + C_{x3} + C_{x5}) + 3(C_{u2} + C_{u4} + C_{u6})] \\
b''_{z2} = \frac{1}{432}[(25C_{x2} + C_{x4} + C_{x6}) + 3(C_{u1} + C_{u4} + C_{u5})] \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
b''_{z6} = \frac{1}{432}[(25C_{u6} + C_{u2} + C_{u4}) + 3(C_{x1} + C_{x3} + C_{x5})].
\end{array} \tag{9}$$

The values of b'''s calculated as above contain the replicate effects. To correct for this, adjust the values so that the sum of the block effects for each replicate is zero. From the corrected b'' values and the normal equations with the v'''s as leading terms, the corrected varietal sums are calculated. These are:

$$4v''_j + 4m'' = V_j - \Sigma b''_{ei} \text{ (sum over blocks in which } v_j \text{ appears)} \tag{10}$$

where $4v''_j + 4m''$ is the corrected varietal total for the jth variety.

3. Test of significance and the analysis of variance. If the x's have the previously defined values, the following identity occurs:

$$Sy^2 = mG + \sum_{e=1}^{8} r_e R_e + \sum_{e=1}^{8}\sum_{i=1}^{6} \frac{(b_{ei} - b'_{ei})}{2}(B_{ei} - B'_{ei})$$

$$+ \sum_{e=1}^{8}\sum_{i=1}^{6} \frac{(b_{ei} + b'_{ei})}{2}(B_{ei} + B'_{ei}) + \sum_{j=1}^{36} v_j V_j \tag{11}$$

$$+ S\left(y - mx_1 - r_e x_{2e} - \frac{b_{ei} - b'_{ei}}{2} x_{3i} - \frac{b_{ei} + b'_{ei}}{2} x_{4i} - v_j x_{5j}\right)^2.$$

In the equation (11)

$$mG + \sum_{e=1}^{8} r_e R_e + \sum_{e=1}^{8}\sum_{i=1}^{6} \frac{(b_{ei} - b'_{ei})}{2}(B_{ei} - B'_{ei}) \tag{12}$$

$$+ \sum_{e=1}^{8}\sum_{i=1}^{6} \frac{(b_{ei} + b'_{ei})}{2}(B_{ei} + B'_{ei}) + \sum_{j=1}^{36} v_j V_j,$$

is the reduction in the sum of squares due to regression and

$$S\left(y - mx_1 - r_e x_{2e} - \frac{b_{ei} - b'_{ei}}{2} x_{3i} - \frac{(b_{ei} + b'_{ei})}{2} x_{4i} - v_j x_5\right)^2, \tag{13}$$

is the residual sum of squares. The reductions attributable to

$$\sum_{e=1}^{8}\sum_{i=1}^{6} \frac{(b_{ei} - b'_{ei})}{2}(B_{ei} - B'_{ei})$$

and

$$\sum_{e=1}^{8}\sum_{i=1}^{6} \frac{(b_{ei} + b'_{ei})}{2}(B_{ei} + B'_{ei}), \tag{14}$$

will be designated as component (a) and component (b) respectively. The residual mean square will be denoted by s^2.

The common test required is that of the null hypothesis that the variety effects $v_1, v_2, \cdots, v_{36}$ are all zero. This test is made by calculating the reduction (R_t) to the sum of squares on all variates, and the reduction (R_v) due to the regression on all variates, except the variety effects. $R_t - R_v$ is called the additional reduction to the sum of squares due to the v's after fitting the remaining variates.

The ratio $(R_t - R_v)/(82 - 47)s^2$ is distributed as F, as shown by Yates (6), with 35 and 205 degrees of freedom. The 35, 205, 82, and 47 degrees of freedom pertain respectively to varieties, error, all constants fitted, and the total constants less the constants for varieties.

Referring to formula (11) and the parameter effects, the sum of squares in the "Analysis of Variance Table" follow directly for replicate and component (a). Nair[2] in his recent article gives in detail the method for getting out the reduction

to the sum of squares for the entangled components. He shows that the reductions for component (b) and the varieties may be written as:

$$(15) \qquad \frac{1}{8}\sum_{e=x}^{u}\sum_{i=1}^{6} b''_{ei}C_{ei} \quad \text{and} \quad \sum_{j=1}^{36}\frac{V_j^2}{8} - \frac{G^2}{N}$$

where the b'' have been corrected for replicate effects. It is well to note that $\frac{1}{8}\sum_{e=x}^{u}\sum_{i=1}^{6} b''_{ei}C_{ei}$ is the reduction due to intra-block effects, freed of varietal effects.

The reduction due to varieties corrected for block effects is given by

$$(16) \qquad \frac{1}{8}\sum_{e=x}^{u}\sum_{i=1}^{6} b''_{ei}C_{ei} + \left(\sum_{j=1}^{36}\frac{V_j^2}{8} - \frac{G^2}{N}\right) - \left(\sum_{e=x}^{u}\sum_{i=1}^{6}\frac{B_{ei}}{6} - \frac{G^2}{N}\right).$$

This reduction can be used for testing the variety effects.

ANALYSIS OF VARIANCE TABLE

	D/F	
Replicate	(8–1)	$\sum_{e=1}^{8}\frac{R_e^2}{36} - \frac{G^2}{N}$
Component (a)	4(6–1)	$\frac{1}{8}\sum_{e=1}^{8}\sum_{i=1}^{6}\frac{(B_{ei}-B'_{ei})^2}{12} - \sum_{e=1}^{8}\frac{(R_e-R'_e)^2}{72}$
Component (b)	4(6–1)	$\frac{1}{8}\sum_{e=x}^{u}\sum_{i=1}^{6} b''_{ei}C_{ei}$
Varieties (ignoring blocks)	(36–1)	$\sum_{j=1}^{36}\frac{V_j^2}{8} - \frac{G^2}{N}$
Error	205	obtained by subtraction
Total	8(36)–1	$Sy^2 - \frac{G^2}{N}$

4. Standard error of adjusted varietal means. For obtaining the standard error of the difference between two varieties adjusted by the intra-block information, this difference between two varieties can be expressed as a linear function of the plot yields. The standard error of the difference then can be obtained from the standard error of a linear function. To obtain the coefficients, it is well to draw a sketch of the plots, and put the coefficient of each plot on the diagram. In this way the proper multipliers can be found in a convenient manner.

First consider the case where the two varieties appear together in the same block in both groups Z and U. One such pair consists of varieties (1) and (22), for which the varietal effects are designated by v''_1 and v''_2. From equation (10) we have:

$$(17) \qquad \begin{aligned} 4v''_1 &= V_1 - 4m'' - b''_{x1} - b''_{y1} - b''_{z1} - b''_{u2} \\ 4v''_{22} &= V_{22} - 4m'' - b''_{x4} - b''_{y4} - b''_{z1} - b''_{u2}\,. \end{aligned}$$

The linear function of the difference between the varietal effects is:

$$4(v''_{22} - v''_1) = V_{22} - V_1 + b''_{x1} - b''_{x4} + b''_{y1} - b''_{y4} \tag{18}$$

where

$$b''_{x1} = \frac{1}{3k}(4B_{x1} - T_{x1}) \qquad b''_{x4} = \frac{1}{3k}(4B_{x4} - T_{x4})$$

$$b''_{y1} = \frac{1}{3k}(4B_{y1} - T_{y1}) \qquad b''_{y4} = \frac{1}{3k}(4B_{y4} - T_{y4}).$$

The multipliers [except for the common factor 4 shown on the left of equation (18)] are:

Number of Plots	*Multipliers*	*Contribution to variance*
4	$\pm\frac{3k+2}{3k}$	$\frac{36k^2+48k+16}{W(9k^2)}$
4	$\pm\frac{3k-2}{3k}$	$\frac{36k^2-48k+16}{W(9k^2)}$
4	$\pm\frac{4}{3k}$	$\frac{64}{W(9k^2)}$
$4(k+2)$	$\pm\frac{3}{3k}$	$\frac{36(k-2)}{W(9k^2)}$
$12(k+2)$	$\pm\frac{1}{3k}$	$\frac{12(k-2)}{W(9k^2)}$
Total		$\frac{72k^2+48k}{W(9k^2)}$

The variance per plot of the difference between two varietal means for varieties which occur together in the same block in groups Z and U is:

$$\frac{72k^2+48k}{2W(16)(9k)} = \frac{3k+2}{12Wk} \tag{19}$$

and for $k = 6$ is $\frac{5}{18W}$.

Similarly the variance per plot of the difference between varietal means for other combinations are as follows:

Combinations	*formula for* $k = 6$
both in same block in groups Z *or* U	$\frac{7}{24W}$
both in same block in groups X *or* Y	$\frac{8}{27W}$
not together in the same block	$\frac{67}{216W}$

5. Numerical Analysis. (a) The data. In order to illustrate the application of the method developed, an experiment used to test the yields of 36 hybrid corn varieties is presented. This experiment was carried out on the Arlington Experimental Farm, and the results are used through the courtesy of A. E. Brandt[4] and M. H. Jenkins.[5]

Except for randomization, the plot yields are as shown in tables I to IV.

(b). *Calculation for analysis of variance table.* From page 9, the total sum of squares, the sum of squares for replicates, and the sum of squares for varieties ignoring blocks are obtained by substitution. They are:

$$(90.9)^2 + (81.4)^2 + \cdots + (101.0)^2 - c = 33{,}546.92$$

$$\frac{(3291.8)^2 + (3300.3)^2 + \cdots + (2978.2)^2}{36} - c = 2{,}289.68$$

and

$$\frac{(741.2)^2 + (695.6)^2 + \cdots + (743.1)^2}{8} - c = 15{,}825.09$$

where

$$c = \frac{(25{,}935.9)^2}{288} = 2{,}335{,}662.80$$

The block sum of squares, eliminating varieties, is made up of two parts, component (a) and component (b). From the formula on page 9 the reduction for component (a) is:

$$\frac{(559.2 - 540.2)^2 + (547.0 - 522.4)^2 + \cdots + (515.8 - 507.7)^2}{12}$$

$$- \frac{(3291.8 - 3300.2)^2 + (3256.5 - 3304.7)^2 + \cdots + (3284.6 - 2978.2)^2}{72} = 1{,}415.27.$$

Component (b) consists of differences giving an estimate of block yield freed of varietal effects. The C's are first calculated by using formula (8) and the results are as follows:

$$C_{x1} = (4B_{x1} - T_{x1}) = 4(1{,}099.4) - 4203.2 = 194.4 \text{ etc.}$$

The b's are calculated by using formulas given by (9), and then correcting for replicate effects by imposing the conditions that

$$\sum_{i=1}^{6} b''_{ei} = 0 \qquad (e = x \cdots u)$$

The corrected b''_{ei} are:

$$b''_1 = 6.79556, \quad b''_2 = -3.83777, \cdots \quad b''_{24} = -0.34306.$$

[4] Acting Chief, Conservation Experiment Station Divisions, Office of Research Soil Conservation Service, U.S.D.A.

[5] Principal Agronomist. In charge of Corn Investigations, Bureau of Plant Industry, U.S.D.A.

TABLE I

GROUP X

Plot Yield Replicate 1

							Σ
	1 90.9	2 81.4	3 95.4	4 96.9	5 96.6	6 98.0	559.2
	7 96.0	8 92.9	9 86.3	10 96.4	11 84.5	12 90.9	547.0
	13 79.3	14 97.5	15 95.9	16 80.2	17 101.8	18 112.1	566.8
	19 83.4	20 92.6	21 91.2	22 88.3	23 71.1	24 111.8	538.4
	25 93.7	26 86.1	27 68.0	28 109.2	29 94.9	30 93.3	545.2
	31 85.3	32 81.2	33 99.5	34 86.9	35 93.4	36 88.9	535.2
Σ	528.6	531.7	536.3	557.9	542.3	595.0	3291.8

Plot Yield Replicate 1′

							Σ
	1 105.8	2 83.5	3 68.6	4 99.0	5 91.8	6 91.5	540.2
	7 98.6	8 71.9	9 81.0	10 98.6	11 90.3	12 82.0	522.4
	13 70.5	14 90.4	15 86.2	16 88.8	17 99.7	18 113.5	549.1
	19 94.6	20 89.6	21 106.8	22 99.2	23 73.2	24 90.9	554.3
	25 91.6	26 86.9	27 74.6	28 104.1	29 102.3	30 96.0	555.5
	31 91.1	32 95.1	33 98.8	34 95.5	35 88.1	36 110.2	578.8
Σ	552.2	516.4	516.0	585.2	545.4	584.1	3300.3

TABLE II

GROUP Y

Plot Yield Replicate 2

						Σ
1 85.9	7 96.6	13 85.4	19 86.4	25 88.8	31 89.7	532.8
2 87.1	8 83.5	14 86.9	20 88.1	26 89.1	32 102.7	537.4
3 88.4	9 85.5	15 90.4	21 99.7	27 80.5	33 100.2	544.7
4 88.5	10 72.3	16 89.5	22 95.6	28 105.5	34 82.6	534.0
5 87.0	11 88.1	17 101.8	23 82.5	29 86.8	35 100.1	546.3
6 90.8	12 83.4	18 101.2	24 120.2	30 72.6	36 93.1	561.3
						3256.5

Plot Yield Replicate 2'

						Σ
1 95.1	7 96.0	13 85.4	19 74.3	25 93.0	31 86.7	530.5
2 96.5	8 84.2	14 79.0	20 88.1	26 95.6	32 101.3	544.7
3 95.4	9 83.3	15 95.9	21 99.0	27 66.0	33 105.3	544.9
4 81.4	10 95.0	16 91.6	22 101.4	28 108.5	34 90.5	568.4
5 91.8	11 107.6	17 84.2	23 66.1	29 87.6	35 101.6	538.9
6 92.2	12 89.6	18 104.1	24 116.0	30 81.5	36 93.9	577.3
						3304.7

TABLE III

GROUP Z

Plot Yield Replicate 3

						Σ
1 87.3	8 79.9	15 86.9	22 105.0	29 99.4	36 90.3	548.8
2 90.0	9 93.0	16 93.1	23 90.3	30 73.9	31 94.1	534.4
3 91.2	10 92.9	17 90.6	24 92.3	25 76.0	32 97.2	540.2
4 87.8	11 98.2	18 125.1	19 102.1	26 88.3	33 98.1	599.6
5 84.3	12 93.7	13 73.9	20 84.4	27 66.7	34 89.8	492.8
6 90.1	7 91.3	14 85.4	21 95.4	28 93.7	35 101.6	557.5
						3273.3

Plot Yield Replicate 3'

						Σ
1 99.4	8 83.5	15 70.9	22 99.9	29 103.0	36 69.0	525.7
2 92.2	9 96.0	16 98.1	23 74.7	30 89.8	31 103.6	554.4
3 91.9	10 106.4	17 92.7	24 108.3	25 88.8	32 103.4	591.5
4 99.0	11 105.4	18 101.2	19 78.1	26 90.5	33 98.1	572.3
5 63.7	12 94.4	13 80.7	20 78.4	27 71.3	34 86.9	475.4
6 81.5	7 88.6	14 90.4	21 84.1	28 93.7	35 88.9	527.2
						3246.5

TABLE IV

GROUP U

Plot Yield Replicate 4

						Σ
31 98.5	26 99.3	21 97.6	16 90.9	11 102.5	6 91.5	580.3
1 95.9	32 92.3	27 70.0	22 99.2	17 85.6	12 100.6	543.6
7 77.8	2 79.9	33 98.1	28 104.8	23 88.9	18 115.7	565.2
13 78.6	8 81.3	3 84.1	34 86.9	29 91.3	24 111.8	534.0
19 82.6	14 89.0	9 75.0	4 94.1	35 108.3	30 96.7	545.7
25 87.3	20 76.2	15 89.7	10 83.7	5 82.2	36 96.7	515.8
						3284.6

Plot Yield Replicate 4'

						Σ
31 83.1	26 88.3	21 84.8	16 94.5	11 73.6	6 84.4	508.7
1 80.9	32 82.6	27 73.9	22 94.8	17 75.8	12 83.4	491.4
7 93.3	2 85.0	33 100.2	28 87.8	23 83.2	18 94.7	544.2
13 74.6	8 68.2	3 45.2	34 79.0	29 87.6	24 102.1	456.7
19 69.1	14 81.2	9 80.3	4 68.8	35 83.7	30 86.4	469.5
25 89.5	20 71.7	15 75.1	10 97.1	5 73.3	36 101.0	507.7
						2978.2

To get the reduction due to component (b) the above results are substituted in

$$\frac{1}{8}\sum_{e=x}^{u}\sum_{i=1}^{6} b''_{ei}\,C_{ei} = 1{,}389.96.$$

The necessary results are now available for the "Analysis of Variance Table."

THE ANALYSIS OF VARIANCE TABLE

Source of Variation	*Degrees of Freedom*	*Sum of Squares*	*Mean Square*
Replications	7	2,289.68	327.097
Component (a)	20	1,415.27	
Component (b)	20	1,389.96	
Blocks (eliminating varieties)	40	2,805.23	70.131
Varieties (ignoring blocks)	35	15,825.09	
Error	205	12,626.92	61.595
Total	287	33,546.92	

(c). *Test of significance.* There frequently will be large differences between varieties so that a test of significance may not be needed. If a test is needed, one involving only intra-block information may be used. For this purpose, it is necessary to calculate the sum of squares for varieties eliminating block effects as shown by formula (15): 13,946.28. The mean square will be 399.893, and $F = \dfrac{399.893}{61.595} = 6.49$ which is highly significant.

(d). *Corrected varietal totals and means.* The right-hand side of equation (10) gives the corrected variety totals, and when divided by eight gives the varietal means. These corrected varietal means can then be compared to determine the best variety. The corrected varietal totals and means are:

Corrected Varietal Totals

1 743.22	2 669.05	3 652.00	4 705.80	5 672.04	6 720.32
7 747.59	8 658.58	9 664.57	10 751.39	11 739.54	12 735.95
13 642.31	14 700.44	15 686.41	16 713.21	17 730.79	18 857.26
19 665.95	20 675.40	21 756.25	22 801.34	23 619.46	24 868.84
25 704.17	26 699.83	27 567.71	28 814.04	29 763.51	30 679.44
31 721.79	32 757.48	33 783.42	34 726.05	35 780.48	36 760.37

Corrected Varietal Means

1 92.902	2 83.631	3 81.500	4 88.225	5 84.005	6 90.040
7 93.449	8 82.322	9 83.071	10 93.924	11 92.442	12 91.994
13 80.289	14 87.555	15 85.801	16 89.151	17 91.349	18 107.158
19 83.244	20 84.425	21 94.531	22 100.168	23 77.432	24 108.605
25 88.021	26 87.479	27 70.964	28 101.755	29 95.439	30 84.930
31 90.224	32 94.685	33 97.927	34 90.756	35 97.560	36 95.046

When comparing one variety with another it is necessary to know the standard error of the mean difference, in order to judge whether this difference is significant. The formulas for the standard error of the difference between mean yields differ for those sets of varieties which occur together in the same block in groups Z and U, in groups Z or U, in groups X or Y, or do not occur together in the same block. The formulas for these calculations are, respectively:

(19), (20), (21), and (22). For example, the standard error of the difference between two variety means, such as variety 1 and 2, is

$$\frac{8}{27}(61.595) = 4.27.$$

REFERENCES

[1] G. M. Cox, R. C. Eckhardt and W. G. Cochran, "The analysis of lattice and triple lattice experiments in corn varietal tests," *Iowa Agr. Exp. Sta. Res. Bul.*, Vol. 281 (1940).

[2] K. P. Nair, "A note on the method of fitting constants," *Sankhya*, Vol. 5 (1941), pp. 317–328.

[3] F. Yates, "The principles of orthogonality and confounding in replicated experiments," *Journal Agr. Sci.*, Vol. 23 (1933), 108–145.

[4] F. Yates, "Complex experiments," *Suppl. Jour. Roy. Stat. Soc.*, Vol. 2 (1935), pp. 181–247.

[5] F. Yates, "A new method of arranging variety trials involving a large number of varieties," *Journal Agr. Sci.*, Vol. 26 (1936), pp. 424–455.

[6] F. Yates, "Orthogonal functions and tests of significance in the analysis of variance," *Journal Royal Stat. Soc.*, Vol. 5 (1938), pp. 177–180.

NOTES

This section is devoted to brief research and expository articles on methodology and other short items.

ON THE EXPECTED VALUES OF TWO STATISTICS

By H. E. Robbins

Post Graduate School; Annapolis, Md.

In a previous paper[1], the following theorem was proved. Let X be a random, Lebesgue measurable subset of Euclidean m dimensional space E_m, and let $\mu(X)$ be the measure of X. For every point x of E_m let $p(x)$ be the probability that X contains x. Then

$$E(\mu(X)) = \int_{E_m} p(x)\, d\mu(x). \tag{1}$$

In the present note we shall show how this theorem may be used to find the expected values of two statistics which arise in sampling theory. Applications to similar problems may suggest themselves to the reader.

Let Y be a real random variable with c. d. f. (cumulative distribution function) $\sigma(y)$, so that for every y,

$$\Pr\ (Y < y) = \sigma(y). \tag{2}$$

Let $Y_1, \cdots, Y_n$ be n independent random variables, each with the distribution of Y. Finally, let

$$\begin{aligned} A &= \min\ (Y_1, \cdots, Y_n), \\ B &= \max\ (Y_1, \cdots, Y_n), \\ R &= B - A, \\ F &= \sigma(B) - \sigma(A). \end{aligned} \tag{3}$$

Although the values of $E(F)$ and $E(R)$ can be found from the sampling distributions of F and R, and, in fact, are well known, we shall show how to apply (1) to find $E(F)$ and $E(R)$ directly.

To find the first of these, let X denote the set of points in the interval $0 \leq x \leq 1$ such that

$$\sigma(A) < x < \sigma(B). \tag{4}$$

Then X is a random set with measure

$$\mu(X) = F. \tag{5}$$

Moreover, for any point x the probability that X shall contain x is clearly

$$p(x) = 1 - x^n - (1 - x)^n. \tag{6}$$

Hence by (1),

$$E(\mu(X)) = \int_0^1 [1 - x^n - (1 - x)^n]\, dx = \frac{n-1}{n+2}. \tag{7}$$

Thus by (5),

$$E(F) = \frac{n-1}{n+1}. \tag{8}$$

This result may also be derived by the usual method. In fact, it is not hard to show that the c. d. f. of F is

$$\tau(f) = \Pr(F < f) = (1 - n)f^n + nf^{n-1} \quad \text{for} \quad 0 \leq f \leq 1, \tag{9}$$

whence

$$\begin{aligned} E(F) &= \int_0^1 f\, d\tau(f) = (1 - n)n \int_0^1 f^n\, df + n(n-1) \int_0^1 f^{n-1}\, df \\ &= \frac{(1-n)n}{n+1} + \frac{n(n-1)}{n} = \frac{n-1}{n+1}. \end{aligned} \tag{10}$$

Here the advantage of using (1) is only that it makes unnecessary the calculation of the c. d. f. $\tau(f)$, provided that only $E(F)$ is desired.

The situation is quite otherwise with $E(R)$. Here the c. d. f. of R is

$$\theta(r) = \Pr(R < r) = n(n-1) \int_{-\infty}^{\infty} \varphi(a) \int_a^{a+r} \varphi(b) \left[\int_a^b \varphi(t)\, dt \right]^{n-2} db\, da, \tag{11}$$

where φ is the probability density function of Y. Unless φ is a very simple function, it would seem difficult to find a simple expression for the integral

$$E(R) = \int_0^{\infty} r\, d\theta(r). \tag{12}$$

However, if we let X now denote the linear set

$$A \leq t \leq B, \tag{13}$$

then

$$\mu(X) = B - A = R. \tag{14}$$

The probability that X shall contain the point t is now

$$p(t) = 1 - \sigma^n(t) - (1 - \sigma(t))^n, \tag{15}$$

so that, by (1) and (14),

$$E(R) = \int_{-\infty}^{\infty} \{1 - \sigma^n(t) - (1 - \sigma(t))^n\}\, dt. \tag{16}$$

This formula for the expected value of the range in a sample of n from a population Y with c. d. f. $\sigma(t)$ is believed to be new.

If $\sigma(t)$ is such that $dt/d\sigma$ can be found as an explicit function of σ, then (16) can be written with advantage as

$$E(R) = \int_0^1 \{1 - \sigma^n - (1 - \sigma)^n\} \frac{dt}{d\sigma} d\sigma. \tag{17}$$

For example, suppose the random variable Y has the probability density function

$$\varphi(y) = \frac{e^y}{(1 + e^y)^2}, \tag{18}$$

and hence the c. d. f.

$$\sigma(y) = \frac{e^y}{1 + e^y}. \tag{19}$$

Then

$$t = \log \frac{\sigma}{1 - \sigma}, \qquad \frac{dt}{d\sigma} = \frac{1}{\sigma(1 - \sigma)}. \tag{20}$$

Hence from (17), the expected value of the range in a sample of n is

$$E(R) = \int_0^1 \frac{1 - \sigma^n - (1 - \sigma)^n}{\sigma(1 - \sigma)} d\sigma. \tag{21}$$

The indicated division in the integrand may be carried out, and the result, a polynomial in σ of degree $\leq (n - 2)$, when integrated between 0 and 1, gives an explicit formula for $E(R)$. Thus for samples of $n = 2, 3, 4$ we find the values of $E(R)$ to be respectively 2, 3, 11/3. Incidentally, it is always true that the expected value of the range for $n = 3$ is three-halves that for $n = 2$. This follows from (16) and the algebraic identity

$$\{1 - \sigma^3 - (1 - \sigma)^3\} = \tfrac{3}{2}\{1 - \sigma^2 - (1 - \sigma)^2\}. \tag{22}$$

REFERENCES

[1] H. E. Robbins, "On the measure of a random set," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 70–74.

ON RELATIVE ERRORS IN SYSTEMS OF LINEAR EQUATIONS

By A. T. Lonseth

Northwestern University

Some time ago in these *Annals*[1], L. B. Tuckerman discussed the effect of relative coefficient errors on relative solution errors for a non-singular linear algebraic system; his discussion was confined to errors so small that their squares and higher powers can be neglected. Dr. Tuckerman's paper was principally concerned

[1] L. B. Tuckerman, "On the mathematically significant figures in the solution of simultaneous linear equations," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 307–316.

with the important problem of limiting errors incurred while solving the system, and has suggested the desirability of a non-infinitesimal treatment of relative errors. Such a treatment follows; the method is a variant of that used in a paper on absolute errors[2]. The computations provide (1) a criterion for allowable relative errors in the coefficients to assure non-vanishing of the determinant; (2) a bound (subject to this criterion) for the relative error in each solution; (3) a specification of accuracy in the coefficients to assure a pre-assigned accuracy in the solution.

We consider a system of linear equations

$$\sum_{j=1}^{n} a_{ij}x_j = c_i \qquad i = 1, 2, \cdots, n, \tag{1}$$

where none of the following vanish: the $n(n+1)$ coefficients a_{ij}, c_i; the determinant Δ of the system; and the n solution-components x_i. Under these conditions it is possible to speak of "relative errors" in the a's, c's, and x's. Let ϵ_{ij}, σ_i be the relative errors in a_{ij} and c_i respectively, so that a_{ij} is perturbed to $a_{ij}(1+\epsilon_{ij})$; c_i to $c_i(1+\sigma_i)$. We inquire as to limitations on ϵ_{ij} and σ_i which will permit solution for $x_j(1+\rho_j)$ of the system

$$\sum_{j=1}^{n} a_{ij}(1+\epsilon_{ij})x_j(1+\rho_j) = c_i(1+\sigma_i) \qquad i = 1, 2, \cdots, n, \tag{2}$$

where ρ_j is the relative error induced in x_j; and we seek to limit $|\rho_j|$ in terms of the ϵ's and σ's. We shall assume that for all i, j

$$|\epsilon_{ij}|, |\sigma_i| < \delta, \tag{3}$$

where δ will be suitably restricted later.

Combining (1) and (2) we get

$$\sum_{j=1}^{n} a_{ij}x_j\rho_j = c_i\sigma_i - \sum_{j=1}^{n} a_{ij}\epsilon_{ij}x_j - \sum_{j=1}^{n} a_{ij}\epsilon_{ij}x_j\rho_j \qquad i = 1, 2, \cdots, n, \tag{4}$$

Since by hypothesis the determinant Δ of (1) is not zero, matrix $A = (a_{ij})$ has the inverse $A^{-1} = (b_{ij}) = (A_{ji}/\Delta)$, where A_{ij} is the cofactor of a_{ij} in Δ. Multiplying (4) by b_{ki} and summing on i we get

$$x_k\rho_k = \sum_{i=1}^{n} b_{ki}c_i\sigma_i - \sum_{i=1}^{n} b_{ki}\sum_{j=1}^{n} a_{ij}\epsilon_{ij}x_j - \sum_{i=1}^{n} b_{ki}\sum_{j=1}^{n} a_{ij}\epsilon_{ij}x_j\rho_j \qquad k = 1, 2, \cdots, n, \tag{5}$$

and by (3)

$$\rho \leqq \delta(M_k + N_k|\rho_k|), \tag{6}$$

[2] A. T. Lonseth, "Systems of linear equations with coefficients subject to error," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 332–337.

where ρ is the greatest $|\rho_k|$, and

$$M_k = \frac{1}{|x_k|}\sum_{i=1}^{n}|b_{ki}|\left(|c_i| + \sum_{j=1}^{n}|a_{ij}x_j|\right), \tag{7}$$

$$N_k = \frac{1}{|x_k|}\sum_{i=1}^{n}\sum_{j=1}^{n}|b_{ki}a_{ij}x_j|, \tag{8}$$

so that

$$M_k = N_k + \frac{1}{|x_k|}\sum_{i=1}^{n}|b_{ki}c_i|.$$

Denote by M, N the maximum values of M_k, N_k respectively over $k = 1, 2, \cdots, n$. From (6),

$$\rho \leqq \delta(M + N\rho),$$

whence, if

$$\delta < 1/N, \tag{9}$$

it follows that

$$\rho \leqq \delta M/(1 - \delta N), \tag{10}$$

which of course bounds each individual $|\rho_k|$, though rather crudely. To bound $|\rho_k|$ more genuinely in terms of δ, M_k, N_k, M and N it remains only to use this inequality with (6):

$$|\rho_k| \leqq \delta(M_k + \delta MN_k/(1 - \delta N)), \qquad k = 1, 2, \cdots, n; \tag{11}$$

with M_k, N_k as given in (7) and (8).

If (9) holds, then, it follows that $|\rho_k|$ is bounded by (11)—if ρ_k exists. This essential point can be established by solving (5) for ρ_k by iteration[2]; (9) is a sufficient condition for convergence of the resulting series, and hence for non-singularity of the perturbed matrix $(a_{ij} + \epsilon_{ij}a_{ij})$.

In order to be sure that $|\rho_k| \leqq \eta$, a pre-assigned number, for all k, it suffices by (10) to choose δ so that

$$\delta M/(1 - \delta N) \leqq \eta,$$

whence

$$\delta \leqq \eta/(M + N\eta).$$

A less simple inequality whose satisfaction by δ will guarantee that $|\rho_k| \leqq \eta_k$ follows from (11), namely

$$\delta \leqq (A - B)/2C,$$

where $A = \{(M_k - N\eta_k)^2 + 4MN_k\eta_k\}^{\frac{1}{2}}$, $B = M_k + N\eta_k$, $C = |MN_k - NM_k|$.

A RECIPROCITY PRINCIPLE FOR THE NEYMAN-PEARSON THEORY OF TESTING STATISTICAL HYPOTHESES

By Louis M. Court

In contrasting the tested hypothesis H_1 with the alternative H_2, i.e., in comparing the probability distribution $p(x_1, \cdots, x_n \mid H_1)$ associated with the first hypothesis with the distribution $p(x_1, \cdots, x_n \mid H_2)$ associated with the second, Neyman and Pearson select the *best* critical region R^* from the infinite set of critical regions R of a specified size α by minimizing the probability

$$\int_{S-R} p(x_1, \cdots, x_n \mid H_2)\, dx_1 \cdots dx_n \tag{1}$$

of accepting H_1 when H_2 is true (Type II Errors) subject to the constancy of the probability of rejecting H_1 when H_1 is correct (Type I Errors),

$$\int_R p(x_1, \cdots, x_n \mid H_1)\, dx_1 \cdots dx_n = \alpha. \tag{2}$$

S in (1) represents the whole of variate $(x_1, \cdots, x_n)$ space and $S - R$, the complement of R relative to S.

Obviously (1) is conditionally minimized when

$$\int_R p(x_1, \cdots, x_n \mid H_2)\, dx_1 \cdots dx_n \tag{3}$$

is maximized subject to (2). Neyman and Pearson have shown that if one or more members of the one parameter (λ) family of regions $R^*(\lambda)$ defined by the inequalities

$$p(x_1, \cdots, x_n \mid H_2) \geqq \lambda p(x_1, \cdots, x_n \mid H_1) \tag{4}$$

satisfy the "side" condition (2), they will be best critical regions maximizing (3) subject to (2) or minimizing (1) subject to (2)[1]. As suggested by the notation, the family $R^*(\lambda)$ depends upon λ and, if sufficient restrictions are imposed upon $p(x_1, \cdots, x_n \mid H_1)$ and $p(x_1, \cdots, x_n \mid H_2)$, there is *one and only one* region for every value of λ lying in an interval contained in the positive half-axis. λ, itself, is clearly a function $\lambda(\alpha)$ of α. Consequently, $R^*(\lambda)$ depends upon α and may be written as $R^*[\alpha]$. The best critical region for a preassigned size $\bar{\alpha}$ is given by $R^*[\bar{\alpha}]$.

Will we get the *same* best critical region if among the regions T that fix the probability of Type II Errors,

$$\int_{S-T} p(x_1, \cdots, x_n \mid H_2)\, dx_1 \cdots dx_n = 1 - \beta, \tag{5}$$

[1] For a full exposition, see Neyman and Pearson, *Stat. Res. Memoirs*, Vol. 1 (1936).

we find the one that minimizes the integral in (2) with R replaced by T, i.e. if we find the one that minimizes the probability of Type I Errors? We shall call this turnabout of the usual process the *reversed* Neyman-Pearson principle. To discover the answer, we note that $\int_S p(x_1, \cdots, x_n \mid H_2)\, dx_1 \cdots dx_n$ is equal to unity and (5) may be rewritten as

$$\int_T p(x_1, \cdots, x_n \mid H_2)\, dx_1 \cdots dx_n = \beta. \tag{6}$$

The regions that minimize the left side of (2) with R replaced by T subject to (6) are obviously identical with those that maximize the negative of this left side subject to (6) multiplied through by -1. The latter problem is formally identical with the one referred to in the second paragraph of this note and, invoking Neyman and Pearson's result, we conclude that the said conditional maximization is effected by the one parameter (μ) family of regions $T^*(\mu^{-1})$ defined by the inequalities

$$-p(x_1, \cdots, x_n \mid H_1) \geqq -\mu p(x_1, \cdots, x_n \mid H_2) \tag{7}$$

or

$$p(x_1, \cdots, x_n \mid H_2) \geqq \frac{1}{\mu}\, p(x_1, \cdots, x_n \mid H_1).$$

μ^{-1} in $T^*(\mu^{-1})$ denotes the reciprocal of μ. It is clear from (4) and (7) that the families of regions $R^*(\lambda)$ and $T^*(\mu^{-1})$, satisfying the direct and reversed Neyman-Pearson processes, coincide.

As before, μ is some function $\mu(\beta)$ of β. Hence, β is a function $\beta(\mu)$[2] of μ and, accordingly, a function $\beta\left[\frac{1}{\lambda(\alpha)}\right]$ of α. Consequently, if the level at which the probability of Type II Errors in the *reversed* Neyman-Pearson process is held constant is taken equal to $1 - \beta\left[\frac{1}{\lambda(\alpha)}\right]$ in terms of the level α at which the probability of Type I Errors is fixed in the *direct* Neyman-Pearson method, the reversed and direct processes yield the *same* best critical region. *This is the reciprocity principle alluded to in the title of this note in its full completeness.*

[2] $\beta(\mu^{-1})$ will generally be distinct in form from $\alpha(\lambda)$, although the second line in (7) coincides upon the substitution of λ for μ^{-1} with (4), since the integrand in (5) is $p(x_1, \cdots, x_n \mid H_2)$ whereas that in (2), regarded as a constraint in the direct process, is $p(x_1, \cdots, x_n \mid H_1)$.

AN INEQUALITY DUE TO H. HORNICH

By Z. W. Birnbaum and Herbert S. Zuckerman
University of Washington

H. Hornich[1] proved a theorem on the average risk of the sum of equal insurance policies. It seems of interest to note that when translated from its actuarial formulation into the terminology of the calculus of probabilities this theorem becomes an inequality for mean deviations of random variables, and to present it with a concise proof in non-actuarial language.

Let x be a random variable with a symmetrical probability distribution, $D_1 = E(|x|)$ its mean deviation, $x_1, x_2, \cdots, x_n$ independent repetitions of x, and $D_n = E(|x_1 + x_2 + \cdots + x_n|)$ the mean deviation of $x_1 + x_2 + \cdots + x_n$. Then D_n fulfills the inequality

$$D_n \geqq \frac{D_1 n}{2^{n-1}} \binom{n-1}{\left[\frac{n}{2}\right]} \tag{1}$$

where $\left[\frac{n}{2}\right]$ denotes the greatest integer $\leqq \frac{n}{2}$. If the distribution of x is not symmetrical but $E(x) = 0$, the inequality becomes

$$D_n \geqq \frac{D_1 n}{2^n} \binom{n-1}{\left[\frac{n}{2}\right]}. \tag{2}$$

The proof will be given for a continuous random variable but it clearly holds quite generally. If $f(x)$ is the probability density of x, and $E(x) = 0$, then one has

$$D_1 = \int_{-\infty}^{+\infty} |x| f(x)\, dx = 2\int_0^{\infty} x f(x)\, dx. \tag{3}$$

In the expression for D_n, the integration over the entire n-space $(x_1, x_2, \cdots, x_n)$ may be performed by integrating separately over each of the 2^n "octants" which correspond to the different combinations of signs of the coordinates, and thus one obtains, for a symmetrical distribution, the estimates

$$D_n = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} \left|\sum_{i=1}^{n} x_i\right| \prod_{i=1}^{n} f(x_i) \prod_{i=1}^{n} dx_i$$

$$= \sum_{\epsilon_1=\pm 1,\, \epsilon_2=\pm 1,\, \cdots,\, \epsilon_n=\pm 1} \int_{\operatorname{sgn} x_1=\epsilon_1}\int_{\operatorname{sgn} x_2=\epsilon_2}\cdots\int_{\operatorname{sgn} x_n=\epsilon_n}$$

$$= \sum_{s=0}^{n}\binom{n}{s} \int_{\substack{\operatorname{sgn} x_1=\operatorname{sgn} x_2=\cdots=\operatorname{sgn} x_s=-1 \\ \operatorname{sgn} x_{s+1}=\operatorname{sgn} x_{s+2}=\cdots=\operatorname{sgn} x_n=+1}}\int\cdots\int \geqq 2\sum_{s=0}^{[(n-1)/2]}\binom{n}{s} \int_{\substack{\operatorname{sgn} x_1=\operatorname{sgn} x_2=\cdots=\operatorname{sgn} x_s=-1 \\ \operatorname{sgn} x_{s+1}=\operatorname{sgn} x_{s+2}=\cdots=\operatorname{sgn} x_n=+1}}\int\cdots\int$$

[1] Hans Hornich, "Zur theorie des Risikos," *Monatsh. Math. Phys.*, Vol. 50 (1941), pp. 142–150.

$$\geqq 2\sum_{s=0}^{[(n-1)/2]}\binom{n}{s}\int\cdots\int_{\substack{\operatorname{sgn}x_1=\operatorname{sgn}x_2=\cdots=\operatorname{sgn}x_s=-1\\ \operatorname{sgn}x_{s+1}=\operatorname{sgn}x_{s+2}=\cdots=\operatorname{sgn}x_n=+1}}\left(\sum_{i=s+1}^{n}x_i-\sum_{i=1}^{s}x_i\right)\prod_{i=1}^{n}f(x_i)\prod_{i=1}^{n}dx_i$$

$$=2\sum_{s=0}^{[(n-1)/2]}\binom{n}{s}(n-2s)\frac{D_1}{2}\cdot\frac{1}{2^{n-1}}=\frac{D_1}{2^{n-1}}\left\{n\sum_{s=0}^{[(n-1)/2]}\binom{n}{s}-2\sum_{s=0}^{[(n-1)/2]}s\binom{n}{s}\right\}$$

$$=\frac{D_1 n}{2^{n-1}}\left\{\sum_{s=0}^{[(n-1)/2]}\binom{n}{s}-2\sum_{s=0}^{[(n-1)/2-1]}\binom{n-1}{s}\right\}$$

$$=\frac{D_1 n}{2^{n-1}}\left\{\binom{n}{\left[\frac{n-1}{2}\right]}+\sum_{s=0}^{[(n-1)/2-1]}\left(\binom{n}{s}-\binom{n-1}{s}\right)-\sum_{s=0}^{[(n-1)/2-1]}\binom{n-1}{s}\right\}$$

$$=\frac{D_1 n}{2^{n-1}}\left\{\binom{n}{\left[\frac{n-1}{2}\right]}+\sum_{s=1}^{[(n-1)/2-1]}\binom{n-1}{s-1}-\sum_{s=0}^{[(n-1)/2-1]}\binom{n-1}{s}\right\}$$

$$=\frac{D_1 n}{2^{n-1}}\left\{\binom{n}{\left[\frac{n-1}{2}\right]}-\binom{n-1}{\left[\frac{n-1}{2}-1\right]}\right\}$$

$$=\frac{D_1 n}{2^{n-1}}\binom{n-1}{\left[\frac{n-1}{2}\right]}=\frac{D_1 n}{2^{n-1}}\binom{n-1}{\left[\frac{n}{2}\right]}.$$

If x is not symmetrical but $E(x)=0$, we consider the random variable x' with the probability density $g(x')=f(-x')$, and the random variable $y=x+x'$. In view of (3) we find

$$E(|y|)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}|x+x'|f(x)g(x')\,dx\,dx'\geqq 2\int_0^{\infty}\int_0^{\infty}(x+x')f(x)g(x')\,dx\,dx'$$

$$=2\int_0^{\infty}g(x')\int_0^{\infty}xf(x)\,dx\,dx'+2\int_0^{\infty}f(x)\int_0^{\infty}x'g(x')\,dx'\,dx$$

$$=E(x)\left\{-\int_0^{\infty}g(-x)\,dx+\int_0^{\infty}f(x)\,dx\right\}=E(x).$$

Let $x_1, x_2, \cdots, x_n$ and $x_1', x_2', \cdots, x_n'$ be independent repetitions of x and x', respectively, and $y_i=x_i+x_i'$. Since y has a symmetrical distribution, an application of (1) gives

$$E(|x_1+x_2+\cdots x_n|)=\tfrac{1}{2}\{E(|x_1+\cdots+x_n|)+E(|x_1'+\cdots+x_n'|)\}$$

$$\geqq\tfrac{1}{2}E(|x_1+x_1'+\cdots+x_n+x_n'|)=\tfrac{1}{2}E(|y_1+\cdots+y_n|)$$

$$\geqq\frac{1}{2}\cdot\frac{E(|y|)n}{2^{n-1}}\binom{n-1}{\left[\frac{n}{2}\right]}\geqq\frac{E(|x|)n}{2^n}\binom{n-1}{\left[\frac{n}{2}\right]}.$$

An application of Stirling's formula shows that the right hand sides in (1) and (2) are of the order of magnitude of $\sqrt{n}$.

NOTE ON A LEMMA

By Abraham Wald

Columbia University

In a previous paper on the power function of the analysis of variance test[1], the author stated the following lemma (designated there as Lemma 2):

Lemma 2. *Let $v_1, \cdots, v_k$ be k normally and independently distributed variates with a common variance σ^2. Denote the mean value of v_i by α_i $(i = 1, \cdots, k)$ and let $f(v_1, \cdots, v_k, \sigma)$ be a function the variables $v_1, \cdots, v_k$ and σ which does not involve the mean values $\alpha_1, \cdots, \alpha_k$. Then, if the expected value of $f(v_1, \cdots, v_k, \sigma)$ is equal to zero, $f(v_1, \cdots, v_k, \sigma)$ is identically equal to zero, except perhaps on a set of measure zero.*

In the paper mentioned above it was intended to state this lemma for bounded functions $f(v_1, \cdots, v_k)$ and the lemma was used there only in a case where $f(v_1, \cdots, v_k)$ is bounded. Through an oversight this restriction on $f(v_1, \cdots, v_k)$ was not stated explicitly.[2] The published proof of Lemma 2 is adequate if $f(v_1, \cdots, v_k)$ is assumed to be bounded. From the fact that the moments of a multivariate normal distribution determine uniquely the distribution it is concluded there that if for any set $(r_1, \cdots, r_k)$ of non-negative integers

$$(1) \qquad \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} v_1^{r_1} \cdots v_k^{r_k} f(v_1, \cdots, v_k) e^{-\frac{1}{2}\Sigma(v_i-\alpha_i)^2}\, dv_1 \cdots dv_k = 0$$

identically in the parameters $\alpha_1, \cdots, \alpha_k$ then $f(v_1, \cdots, v_k)$ must be equal to zero except perhaps on a set of measure zero. This conclusion is obvious if $f(v_1, \cdots, v_k)$ is bounded. In fact, from (1) and the boundedness of $f(v_1, \cdots, v_k)$ it follows that there exists a finite value A such that

$$\varphi(v_1, \cdots, v_k) = \frac{1}{(2\pi)^{k/2}}\left[1 - \frac{1}{A} f(v_1, \cdots, v_k)\right] e^{-\frac{1}{2}\Sigma(v_i-\alpha_i)^2}$$

is a probability density function with moments equal to those of the normal distribution

$$\psi(v_1, \cdots, v_k) = \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}\Sigma(v_i-\alpha_i)^2}.$$

Hence $f(v_1, \cdots, v_k)$ must be equal to zero except perhaps on a set of measure zero. However, this conclusion is not so immediate if no restriction is imposed on $f(v_1, \cdots, v_k)$ except that

$$(2) \qquad \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} |f(v_1, \cdots, v_k)|\, e^{-\frac{1}{2}\Sigma(v_i-\alpha_i)^2}\, dv_1 \cdots dv_k < \infty$$

for all values of the parameters $\alpha_1, \cdots, \alpha_k$. It is the purpose of this note to prove this. In other words, we shall prove the following proposition:

---

[1] A. Wald, "On the power function of the analysis of variance test," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 434.

[2] I wish to thank Prof. J. Neyman for calling my attention to this omission.

PROPOSITION I. *If* (2) *holds for all values of the parameters* $\alpha_1, \cdots, \alpha_k$ *and if for any set* $(r_1, \cdots, r_k)$ *of non-negative integers equation* (1) *holds identically in* $\alpha_1, \cdots, \alpha_k$, *then* $f(v_1, \cdots, v_k)$ *must be equal to zero except perhaps on a set of measure zero.*

On the basis of Proposition I and the arguments given on p. 438 of the paper mentioned before, it can be seen that restriction (2) on the function $f(v_1, \cdots, v_k)$ is sufficient for the validity of Lemma 2.

To prove Proposition I, we shall first show that the following lemma holds.

LEMMA A. *If* $h(v_1, \cdots, v_k)$ *is a probability density function and if*

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} h(v_1, \cdots, v_k) e^{\delta \sum_i |v_i|} \, dv_1 \cdots dv_k < \infty \tag{3}$$

for some $\delta > 0$, *then the problem of moments is determined for the moments of the distribution* $h(v_1, \cdots, v_k)$.

This lemma was proved by G. H. Hardy for $k = 1$.[3] I shall prove it for $k > 1$. Since

$$\sum_{n=0}^{\infty} \frac{\delta^{2n}(\sum |v_i|)^{2n}}{(2n)!} < e^{\delta \Sigma |v_i|} \tag{4}$$

we obtain from (3)

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} h(v_1, \cdots, v_k) \left[\sum_{n=0}^{\infty} \frac{\delta^{2n}(\sum |v_i|)^{2n}}{(2n)!} \right] dv_1 \cdots dv_k < \infty. \tag{5}$$

Hence

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} h(v_1, \cdots, v_k) \left[\sum_{n=0}^{\infty} \frac{\delta^{2n}\left(\sum_{i=1}^{k} v_i^{2n}\right)}{(2n)!} \right] dv_1 \cdots dv_k < \infty. \tag{6}$$

Denote the $2n$th moment of v_i by $\mu_{2n}^{(i)}$. Because of (3) the moments $\mu_{2n}^{(i)}$ are finite. Furthermore, denote $\sum_{i=1}^{k} \mu_{2n}^{(i)}$ by λ_{2n}. Then we obtain from (6)

$$\sum_{n=0}^{\infty} \frac{\delta^{2n}\lambda_{2n}}{(2n)!} < \infty. \tag{7}$$

From (7) it follows that

$$\limsup_{n=\infty} \frac{\delta^{2n}\lambda_{2n}}{(2n)!} < 1. \tag{8}$$

Hence

$$\limsup_{n=\infty} \left(\frac{\delta^{2n}\lambda_{2n}}{(2n)!} \right)^{1/2n} \leq 1. \tag{9}$$

---

[3] See for instance, SHOHAT and TAMARKIN, "The problem of moments," *Math. Surveys No.* 1, *Amer. Math. Soc.*, New York, 1943, p. 20.

Since according to Stirling's formula

$$\lim_{n=\infty} (2n)!/(2n)^{2n}e^{-2n}\sqrt{4\pi n} = 1$$

we obtain from (9)

$$\limsup_{n=\infty} \frac{\delta\lambda_{2n}^{1/2n}}{2ne^{-1}} \leq 1. \tag{10}$$

Taking reciprocals we obtain

$$\liminf_{n=\infty} \frac{2n\lambda_{2n}^{-1/2n}}{e\delta} \geq 1 \tag{11}$$

or

$$\liminf_{n=\infty} n\lambda_{2n}^{-1/2n} \geq \frac{e\delta}{2} > 0. \tag{12}$$

But (12) implies the existence of a positive value ρ so that

$$\lambda_{2n}^{-1/2n} \geq \frac{\rho}{n} \qquad (n = 1, 2, \cdots, \text{ad inf.}) \tag{13}$$

From (13) it follows that

$$\sum_{n=1}^{\infty} \lambda_{2n}^{-1/2n} = \infty. \tag{14}$$

But (14) is Carleman's sufficient condition for the determinateness of the problem of moments. Hence Lemma A is proved.

On the basis of Lemma A we can prove Proposition I as follows: From (2) we obtain

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} |f(v_1, \cdots, v_k)|\, e^{-\frac{1}{2}\Sigma v_i^2 + \Sigma \alpha_i v_i}\, dv_1 \cdots dv_k < \infty \tag{15}$$

for all values $\alpha_1, \cdots, \alpha_k$. Let $f_1(v) = f(v)$ for all points $v = (v_1, \cdots, v_k)$ for which $f(v) \geq 0$, and $f_1(v) = 0$ for all points v for which $f(v) < 0$. Similarly, let $f_2(v) = -f(v)$ for all points v for which $f(v) \leq 0$, and $f_2(v) = 0$ for all points v for which $f(v) > 0$. Then $f_1(v)$ and $f_2(v)$ are non-negative functions and

$$f(v) = f_1(v) - f_2(v). \tag{16}$$

From (15) it follows that

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_1(v) e^{-\frac{1}{2}\Sigma v_i^2 + \Sigma \alpha_i v_i}\, dv_1 \cdots dv_k < \infty \tag{17}$$

and

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_2(v) e^{-\frac{1}{2}\Sigma v_i^2 + \Sigma \alpha_i v_i}\, dv_1 \cdots dv_k < \infty. \tag{18}$$

Let

$$f_j^*(v) = f_j(v)e^{-\frac{1}{2}\Sigma v_i^2} \qquad (j = 1, 2). \tag{19}$$

Now we shall show that for any positive values $\beta_1, \cdots, \beta_k$

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_j^*(v_1, \cdots, v_k)e^{\beta_1|v_1|+\cdots+\beta_k|v_k|}\, dv_1 \cdots dv_k < \infty. \tag{20}$$

In fact, consider the 2^k sets $(a_1, \cdots, a_k)$ where $a_i = \pm 1$ $(i = 1, \cdots, k)$. Denote by $R_{a_1 \cdots a_k}$ the subset of the k-dimensional Cartesian space which consists of all points $v = (v_1, \cdots, v_k)$ for which v_i is either zero or signum $v_i =$ signum a_i $(i = 1, \cdots, k)$. Putting $\alpha_i = a_i\beta_i$, it follows from (17) and (18) that

$$\int_{R_{a_1 \cdots a_k}} f_j^*(v_1, \cdots, v_k)e^{\beta_1|v_1|+\cdots+\beta_k|v_k|}\, dv_1 \cdots dv_k < \infty. \tag{21}$$

Since (21) holds for any of the 2^k sets $R_{a_1 \cdots a_k}$, equation (20) is proved.

From (1) it follows that

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} v_1^{r_1} \cdots v_k^{r_k}[f_1^*(v_1, \cdots, v_k) - f_2^*(v_1, \cdots, v_k)]\, dv_1 \cdots dv_k = 0,$$

for all non-negative integers $r_1, \cdots, r_k$. Hence, because of (21) and Lemma A we see that

$$f_1^*(v_1, \cdots, v_k) = f_2^*(v_1, \cdots, v_k), \tag{22}$$

except perhaps on a set of measure zero. From (22) it follows that

$$f(v_1, \cdots, v_k) = f_1(v_1, \cdots, v_k) - f_2(v_1, \cdots, v_k) = 0,$$

except perhaps on a set of measure zero. Hence Proposition I is proved.

A NOTE ON SKEWNESS AND KURTOSIS

By J. Ernest Wilkins, Jr.

University of Chicago

It is the purpose of §1 of this paper to prove the following inequality:

$$\alpha_4 \geqq \alpha_3^2 + 1. \tag{1}$$

This inequality seems to have first been stated by Pearson[1]. The inequality also follows from a result appearing in the thesis of Vatnsdal. Here we give a proof based on the theory of quadratic forms which seems to be more direct and more elementary than either of the previous proofs.

---

[1] "Mathematical contributions to the theory of evolution, XIX; second supplement to a memoir on skew variation," *Phil. Trans. Roy. Soc.* (*A*), Vol. 216 (1916), p. 432.

The inequality (1) obviously shows that $\alpha_4 \geqq 1$. It is then natural to ask for an upper bound for α_4. In §2 we shall show that there is no universal upper bound (independent of the number N of quantities in the distribution) for α_3. In fact we find the actual dependence of the maximum possible value of α_3 as a function of N. The form of this function seems to be known but not to have been rigorously proved before. It then follows from (1) that there is no universal upper bound for α_4.

1. The inequality (1). Let us consider the quadratic form

$$G(a, b, c) = \nu_0 a^2 + 2\nu_1 ab + 2\nu_2 ac + \nu_2 b^2 + 2\nu_3 bc + \nu_4 c^2$$
$$= N^{-1}\Sigma(a + xb + x^2 c)^2.$$

It follows that $G(a, b, c)$ is a positive semi-definite quadratic form. In fact, if there are at least three distinct values of x, then $G(a, b, c)$ is a positive definite form. Consequently, its discriminant

$$\begin{vmatrix} \nu_0 & \nu_1 & \nu_2 \\ \nu_1 & \nu_2 & \nu_3 \\ \nu_2 & \nu_3 & \nu_4 \end{vmatrix}$$

must be non-negative. There is no loss of generality in supposing that $\nu_1 = 0$, $\nu_2 = 1$, in which case we find that

$$\begin{vmatrix} 1 & 0 & 1 \\ 0 & 1 & \alpha_3 \\ 1 & \alpha_3 & \alpha_4 \end{vmatrix} \geqq 0.$$

Expanding the determinant, we get the inequality (1).

We remark that equality holds in (1) if and only if there are only two distinct values of x.

2. The maximum value of α_3. It is clear that this maximum will be N^{-1} times the maximum value of the function $f(x) = \Sigma x^3$ on the bounded closed set consisting of those points x for which $g(x) = \Sigma x^2 = N$ and $h(x) = \Sigma x = 0$. According to the Lagrange multiplier rule, this latter maximum is obtained as follows. Let $F(x) = f(x) - \lambda g(x) - \mu h(x)$. Then the maximizing point satisfies the relations

$$F_{x_i} = 3x_i^2 - 2\lambda x_i - \mu = 0, \qquad \Sigma x^2 = N, \qquad \Sigma x = 0.$$

The equations $\Sigma F_x = 0$, $\Sigma x F_x = 0$ shows that $\mu = 3$, $f_{\max} = 2N\lambda/3$. Solving the equation $F_{x_i} = 0$ gives

$$x_i = [\lambda + e_i(\lambda^2 + 9)^{\frac{1}{2}}]/3, \tag{2}$$

where $e_i = \pm 1$. For these values of x_i we shall have $h(x) = 0$, $g(x) = N$ if and only if

$$\lambda = -(\lambda^2 + 9)^{\frac{1}{2}} N^{-1} \Sigma e.$$

Therefore λ has the sign opposite to that of Σe, and

$$\lambda^2[N^2 - (\Sigma e)^2] = 9(\Sigma e)^2.$$

It follows that $\Sigma e \neq \pm N$, and that

$$\lambda = -3\Sigma e/[N^2 - (\Sigma e)^2]^{\frac{1}{2}}, \tag{3}$$

$$f_{\max} = -2N\Sigma e/[N^2 - (\Sigma e)^2]^{\frac{1}{2}}.$$

We have still not obtained the maximum, however, since the minimum will also satisfy all of the relations deduced above. We distinguish the maximum from the other critical values by examining the function

$$\theta(\Sigma) = -2N\Sigma/(N^2 - \Sigma^2)^{\frac{1}{2}}.$$

Since $\Sigma e \neq \pm N$, $e_i = \pm 1$, it is clear that $N - 2 \geqq \Sigma e \geqq 2 - N$. We therefore consider $\theta(\Sigma)$ on the interval $(2 - N, N - 2)$. We find that

$$d\theta/d\Sigma = -2N^3/(N^2 - \Sigma^2)^{3/2} < 0,$$

so that θ is a decreasing function of Σ on the interval indicated. Its maximum value will therefore occur when $\Sigma = 2 - N$, and this maximum value will be

$$\theta(2 - N) = N(N - 2)/(N - 1)^{\frac{1}{2}}.$$

The value $\Sigma e = 2 - N$ occurs only when one of the e_i, say e_1, is equal to $+1$ and all the rest are equal to -1. Then we find from (3) and (2) that

$$\lambda = 3(N - 2)/2(N - 1)^{\frac{1}{2}},$$

$$x_1 = (N - 1)^{\frac{1}{2}}, \qquad x_2 = x_3 = \cdots = x_N = -(N - 1)^{-\frac{1}{2}}, \tag{4}$$

$$\alpha_3 = f(x)/N = (N - 2)/(N - 1)^{\frac{1}{2}}.$$

Since the maximum value of α_3 given by (4) approaches ∞ with N, it follows that there is no universal upper bound for α_3. More precisely, the quantity α_3 can be made as large as desired by choosing N large enough and then picking x_i as in the last paragraph. Since there is no universal upper bound for α_3, it is clear from (1) that there is no universal upper bound for α_4. It would probably be possible, although rather difficult, to derive an explicit bound for α_4 as a function of N by using the methods employed above for α_3.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of general interest

Personal Items

Mr. Carl A. Bennett is doing war research on a project at the University of Chicago.

Dr. George W. Brown, formerly Research Associate at Princeton University, is now at the RCA Laboratories in Princeton, New Jersey.

Professor Harry Carver is on leave of absence from the University of Michigan to do Operations Analysis work with the Army Air Forces.

Mr. George B. Dantzig is now Principal Statistician of the Army Air Forces Statistical Control Division at the Pentagon Building in North Arlington, Virginia.

Assistant Professor Preston C. Hammer of Oregon State College has taken a leave of absence to aid in setting up statistical methods of quality control for the Lockheed Aircraft Corporation of Burbank, California.

Dr. Tjalling Koopmans is now associated with the Cowles Commission for Research in Economics at the University of Chicago.

Dr. Jerome C. R. Li is now an instructor in mathematics at Queens College in Flushing, New York.

Associate Professor Joe Livers of Montana State College has been granted a leave of absence for the summer and fall terms to study at the University of Michigan.

Assistant Professor Eugene Lukacs of Illinois College, Jacksonville, Illinois, has accepted an associate professorship in mathematics at Berea College, Berea, Kentucky.

Mr. R. I. Piper, who has been on leave from the Southern California Telephone Company to do war research at the California Institute of Technology, has returned to the Telephone Company as Plant Staff Assistant.

Dr. Henry Scheffé, formerly lecturer in mathematics at Princeton University, has been appointed to an assistant professorship in mathematics at Syracuse University.

Assistant Professor H. M. Schwartz of the University of Idaho has been appointed research fellow at the Bartol Research Foundation in Swarthmore, Pennsylvania.

New Members

The following persons have been elected to membership in the Institute:

Bower, O. K. Ph.D. (Illinois) Asso. in Dept. of Math., Univ. of Illinois. 505 W. John, Champaign, Ill.

Breden, Robert E. B.S. (Kansas State College) Personnel Technician, Statistical Services, Technical Section of the War Dept., 270 Madison Ave., New York, N. Y.

Brixey, John C. Ph.D. (Chicago) Asso. Prof. of Math., Univ. of Oklahoma. 927 S. Pickard, Norman, Okla.

Brown, Arthur B. Ph.D. (Harvard) Asst. Prof. of Math., Queens College, Flushing, N. Y.

Bruyere, Martha (Mrs. Paul T.) Stat., U. S. Public Health Service, Div. of Venereal Diseases, Gaithersburg, Md.

Bruyere, Paul T. M.P.H. (Yale) Stat., U. S. Public Health Service, Bldg. T6, Bethesda, Md.

Carter, Gerald C. Ph.D. (Purdue) Dept. Head, Naval Training School, Purdue Univ. 530 Garfield St., W. Lafayette, Ind.

Casanova, Teobaldo. Ph.D. (New York) Res. Stat., Univ. of Puerto Rico, Rio Piedras, P. R.

Chances, Ralph. B.B.S. (C.C.N.Y.) Stat., Industrial Surveys Co., 347 Madison Ave., New York, N. Y.

Cody, Donald D. A.B. (Harvard) Res. Math., Res. Lab., Indianapolis Naval Ordnance Plant, Indianapolis, Ind.

Duvall, George E. Asst. Physicist, UCDWR, U. S. Navy Radio and Sound Lab., San Diego 52, Calif.

Ellis, Wade. Ph.D. (Michigan) Special Instr. in Math., Univ. of Michigan. 921 Woodlawn, Ann Arbor, Mich.

Field, Robert W. Ph.D. (Illinois) Asso. Prof. of Industrial Engineering, Purdue Univ., Lafayette, Ind.

File, Quentin W. Ph.D. (Purdue) Instr. of Elec. Wiring, Purdue Naval Training School. Physics Bldg., Purdue Univ., Lafayette, Ind.

Freeman, Albert M. Dir. Math. Lab., Boston Fiduciary & Research Associates. Neck Rd., Tiverton, R. I.

Gerlough, Daniel L. B.S. (Calif. Inst. of Tech.) Quality Control Engineer, Plomb Tool Co., Box 3519, Terminal Annex, Los Angeles 54, Calif.

Germond, Hallett H. Ph.D. (Wisconsin) Asso. Prof. of Math., Univ. of Florida, Gainesville, Fla. (On leave).

Ghormley, Glen E. Stat. Analyst, Lockheed Aircraft Corp. 139 N. Chester Ave., Pasadena 4, Calif.

Grant, Eugene L. A.M. (Columbia) Prof. of Economics of Engineering, Stanford Univ., Calif.

Gunlogson, L. S. B.B.A. (Minnesota) Ensign, USNR. 3616 18th Ave., So., Minneapolis 7, Minn.

Hart, Alex L. Ph.D. (Minnesota) Asst. to the Dir., Res. Dept., Eastern Air Lines, Inc. 141-24 79 Ave., Flushing, L. I., N. Y.

Kefferstan, William F. Mgr., Economic Research, Boston Fiduciary & Research Associates, 50 Congress St., Boston, Mass.

Lyons, Will. B.Sc. (Bucknell) Economist, Gen. Statistics Staff, W.P.B. 2027 Park Rd., N.W., Washington 10, D. C.

McBee, Ethelyne L. M.A. (Columbia) Stat., Dept. of Agric. 2126 N. Stafford St., Arlington, Va.

McIntyre, Donald P. M.A. (Toronto) Meteorologist in Charge, Prince George Airport. Box 296, Prince George, B. C., Canada.

Moss, Judith. B.A. (Vassar) Jr. Math., Stat. Res. Group, Div. of War Research, Columbia Univ. 319 St. John's Pl., Brooklyn 17, N. Y.

Oosterhof, Willis M. M.A. (Michigan) Stat., Mich. State Dept. of Social Welfare. 811 Hackett St., Ionia, Mich.

Piper, Robert I. B.A. (Montana) Res. Asso. Room 204, Astrophysics Bldg., Calif. Inst. of Tech., Pasadena 4, Calif.

Robbins, Herbert E. Ph.D. (Harvard) Lt., USNR. Post Graduate School, Annapolis, Md.

Seeley, Sherwood B. Ch.E. (New York) Dir., Res. and Tech. Div., Joseph Dixon Crucible Co., 167 Wayne St., Jersey City 3, N. J.

Simpson, Tracy W. E.E. (Armour Inst. of Tech.) Sales Promotion Mgr., Marchant Calculating Machine Co. 2903 Forest Ave., Berkeley, Calif.

Smith, Edward S. Ph.D. (Virginia) Prof. of Math., Univ. of Cincinnati, Cincinnati 21, Ohio.

Waksberg, Joseph. B.S. (C.C.N.Y.) Economic Analyst, Bureau of the Census. 1422 Saratoga Ave., N.E., Washington 18, D. C.

REPORT ON THE WELLESLEY MEETING OF THE INSTITUTE

The Seventh Summer Meeting of the Institute of Mathematical Statistics was held at Wellesley College, Wellesley, Mass., on Saturday and Sunday, August 12 and 13, 1944, in conjunction with the meetings of the Mathematical Association of America and the American Mathematical Society. The following 51 members of the Institute attended the meeting:

T. W. Anderson, H. E. Arnold, K. J. Arnold, L. A. Aroian, A. L. Bailey, J. L. Barnes, C. I. Bliss, A. H. Bowker, B. H. Camp, C. W. Churchman, W. G. Cochran, T. E. Cope, J. H. Curtiss, W. E. Deming, P. S. Dwyer, Will Feller, C. D. Ferris, R. M. Foster, H. A. Freeman, Henry Goldberg, E. J. Gumbel, P. R. Halmos, Harold Hotelling, Truman Kelley, L. R. Klein, Myra Levine, John Mandel, J. W. Mauchly, Richard v. Mises, E. B. Mode, Vaclav Myslivec, P. M. Neurath, M. L. Norden, C. O. Oakley, P. S. Olmstead, Edward Paulson, Wm. Reitz, S. L. Robinson, F. E. Satterthwaite, Henry Scheffé, W. A. Shewhart, Andrew Sobszyk, H. W. Steinhaus, Marian Torrey, Mary Torrey, A. W. Tucker, J. W. Tukey, Abraham Wald, R. M. Walter, Elizabeth Wilson, Jacob Wolfowitz.

The first session was held jointly with the Mathematical Association and consisted of a Symposium on "Potential Opportunities for Statisticians and the Teaching of Statistics." The President of the Institute, Dr. W. A. Shewhart, presided. The principal addresses were made by Dr. Shewhart and Professor Harold Hotelling. Remarks were also made, upon invitation of the Chairman, by Prof. Milton de Silva Rodrigues of Sao Paulo University in Brazil who is spending a year in studying methods of teaching statistics in this country, and by Dr. Vaclav Myslivec, Czechoslovak Delegate to the United Nations Interim Commission on Food and Agriculture. A lively discussion was under way when time forced the conclusion of the meeting. There was continued discussion by a smaller group for some time afterwards.

Professor B. H. Camp acted as Chairman at the Sunday morning session, a contributed papers session held jointly with the Association. The following papers were presented:

1. *Statistical Tests Based on Permutations of the Observations.*
   A. Wald and J. Wolfowitz, Columbia University.
2. *Error Control in Matrix Calculation.*
   F. E. Satterthwaite, Aetna Life Insurance Co.
3. *On Cumulative Sums of Random Variables.*
   A. Wald, Columbia University.
4. *The Approximate Distribution of the Mean and of the Variance of Independent Variates.*
   P. L. Hsu, National University of Peking. (Introduced, and presented, by W. Feller, Brown University).
5. *Ranges and Midranges.*
   E. J. Gumbel, New School for Social Research.
6. *Statistics of Sensitivity Data, II.* Preliminary report.
   C. W. Churchman, Frankford Arsenal and Benjamin Epstein, Westinghouse Electric and Manufacturing Co.

President Shewhart presided at the Sunday afternoon session. The following invited addresses were given:

1. *The Problem on Tolerance Limits.*
   Lt. J. H. Curtiss, USNR.
2. *Some Improvements in Weighing and Other Experimental Techniques.*
   H. Hotelling, Columbia University.

A business meeting was held at the conclusion of the Sunday afternoon session. The Secretary-Treasurer made a brief report dealing with (1) the financial condition of the Institute and (2) the membership growth of the Institute. The President, reporting for the Editor, indicated a need for more papers for the next two or three issues of the Annals. The Institute, after some discussion, then passed two Amendments to the Constitution and four Amendments to the By-Laws. These Amendments are listed in the following section. A resolution thanking the officials of Wellesley College was passed.

A dinner for the three mathematical organizations was held Sunday evening. Addresses were made by Captain Mildred H. McAfee and Professor Marshall H. Stone. Later in the evening there was a musicale featuring David Barnett.

P. S. DWYER
Secretary

AMENDMENTS TO THE CONSTITUTION AND BY-LAWS OF THE INSTITUTE

The following Amendments to the Constitution and By-laws were passed at the business meeting at Wellesley College on August 13, 1944. The votes of all voting members who sent ballots to the Secretary-Treasurer prior to the time of the meeting were counted in the balloting. The Amendments as adopted are identical with the proposed Amendments which were placed in the hands of the membership in July:

Amendments to Constitution

1. **Article III. 3.** The first sentence, which was

"The Institute shall have a Committee on Membership composed of three Fellows."

shall be revised to read:

"The Institute shall have a Committee on Membership composed of a Chairman and three Fellows."

2. **Article IV. 3.** The first two sentences, which were:

"The Committee on Membership shall hold a meeting immediately after the annual meeting of the Institute. Further meetings of the Committee may be held from time to time at the call of the Chairman or any member of the Committee provided notice of such call and the purpose of the meeting is given to the members of the Committee by the Secretary-Treasurer at least five days before the date set therefor."

shall be revised to read:

"Meetings of the Committee on Membership may be held from time to time at the call of the Chairman or any member of the Committee provided notice of such call and the purpose of the meeting is given to the members of the Committee by the Secretary-Treasurer at least five days before the date set therefor. Committee business may also be transacted by correspondence if that seems preferable.

Amendments to By-Laws

1. **Article I. 4.** Add the following sentence:

"The power of election to the different grades of Membership, except the grades of Member and Junior Member, shall reside in the Board."

2. **Article I. 5.** which was:

"The Committee on Membership shall prepare and make available through the Secretary-Treasurer an announcement indicating the qualifications requisite for the different grades of membership."

shall have added the following sentences:

"The Committee shall review these qualifications periodically and shall make such changes in these qualifications and make such recommendations with reference to the number of grades of membership as it deems advisable. The power to elect worthy applicants to the grades of Member and Junior Member shall reside in the Committee, which may delegate this power to the Secretary-Treasurer, subject to such reservations as the Committee considers appropriate. The Committee shall make recommendations to the Board of Directors with reference to placing members in other grades of membership. The Committee shall give its attention to the question of increasing the number of applicants for membership and shall advise the Secretary-Treasurer on plans for that purpose."

3. After **Article II. 1(a) Exception.** Add:

"(b) Exception. Any Member or Fellow may make a single payment which will be accepted by the Institute in place of all succeeding yearly dues and which will not otherwise alter his status as a Member or Fellow. The amount of this payment will depend upon the age of this Member or Fellow and will be based upon a suitable mortality table and rate of interest, to be specified by the Board of Directors."

and

4. "(c) Exception. Any Member or Junior Member of the Institute serving, except as a commissioned officer, in the Armed Forces of the United States or of one of its allies, may upon notification to the Secretary-Treasurer be excused from the payment of dues until the January first following his discharge from the Service. He shall have all privileges of membership except that he shall not receive the Official Journal. However during the first year of his resumed regular membership he may have the right to purchase, at $2.50 per volume, one copy of each volume of the Official Journal published during the period of his service membership."

ABSTRACTS OF PAPERS

Presented on August 13, 1944, at the Wellesley meeting of the Institute

1. **Statistical Tests Based on Permutations of the Observations.** A. WALD and J. WOLFOWITZ, Columbia University.

It was pointed out by Fisher that statistical tests of exact size, based on permutations of the observations, can be carried out without assuming anything about the underlying distributions except their continuity. Scheffé has proved that, for an important class of hypotheses, these tests are the only ones with regions of exact size. Texts based on permutations of the observations have been constructed by Fisher, Pitman, Welch, and the present authors. In the present paper, the authors prove a theorem on the limiting normality of the distribution, in the universe of permutations, of a class of linear forms. Application of this theorem gives the limiting normality (in the universe of permutations, of course) of the correlation coefficient, and of a statistic introduced by Pitman to test the difference between two means. The limiting distribution of the analysis of variance statistic in the universe of permutations is also obtained.

2. **Error Control in Matrix Calculation.** FRANKLIN E. SATTERTHWAITE, Aetna Life Insurance Co.

The arithmetic evaluation of matrix expressions is often rather complicated. One of the causes of this is the fact that relatively minor errors (such as rounding errors) introduced in an early step may be magnified to such an extent in succeeding steps that the final result is useless. Iterative methods to meet this difficulty have been reviewed very completely by Hotelling. In this paper a different approach is taken. Conditions on the norm of a matrix are determined so that a Doolittle process will not magnify errors to more than two or three decimal places. It is then pointed out that if an approximation to the inverse of the matrix is available, most problems can be rearranged so that the required norm conditions are met. A Doolittle process may then be used to any number of decimal places with assurance that errors will not accumulate to more than a limited number of decimal places.

3. **On Cumulative Sums of Random Variables.** A. WALD, Columbia University.

Let $\{z_i\}$ $(i = 1, 2, \cdots \text{ad inf.})$ be a sequence of independent random variables each having the same distribution. Denote by Z_j the sum of the first j elements of the sequence. Let $a > 0$ and $b < 0$ be two constants and denote by n the smallest integer for which either $Z_n \geq a$ or $Z_n \leq b$. Neglecting the quantity by which Z_n may differ from a or b (this can be done if the mean value of $|z_i|$ is small), the probability that $Z_n \geq c$ for $c = a$ and $c = b$ is derived, and the characteristic function of n is obtained. The probability distribution of n when z_i is normally distributed is derived. These results have application to various statistical problems and to problems in molecular physics dealing with the random walk of particles in the presence of absorbing barriers.

4. **The Approximate Distribution of the Mean and of the Variance of Independent Variates.** P. L. HSU, National University of Peking.

Let X_k be mutually independent random variables with the same cumulative distribution function; let $E(X_k) = 0$, $E(X_k^2) = 1$ and $E(X_k^4) = \delta$. Finally put $S = n^{-1} \sum_{k=1}^{n} X_k$ and $\eta = n^{-1} \sum_{k=1}^{n} (X_k - S)^2$. The author first gives a new derivation of H. Cramer's well-known asymptotic expansions for $Pr\ (n^{\frac{1}{2}}S \leq x)$. The proof is much more elementary and

avoids in particular the use of M. Riesz' singular integrals. Instead a considerably simpler Cesaro-type kernel is used, which has first been introduced by A. C. Berry (*Trans. Amer. Math. Soc.* 49 (1941), pp. 122–136). The same method is then used to derive similar asymptotic expansions for $Pr\ (n^{\frac{1}{2}}(\eta - 1) \leq (\delta - 1)^{\frac{1}{2}}x)$. The method can be extended to the case of unequal components and also for the study of other functions encountered in mathematical statistics.

5. **Ranges and Midranges.** E. J. GUMBEL, New School for Social Research.

The mth range w_m and the mth midrange v_m are defined as the difference and as the sum of the mth extreme value taken in descending magnitude ("from above") and the mth extreme value taken in ascending magnitude ("from below"). The semi-invariant generating functions $L_m(t)$ and ${}_mL(t)$ of the mth extreme values from above and below are simple generalizations of the semi-invariant generating functions of the largest and of the smallest value which have been given by R. A. Fisher and L. H. C. Tippett. If the sample size is large enough the two mth extreme values may be considered as independent variates. Then, the semi invariant generating functions $L_w(t, m)$ and $L_v(t, m)$ of the mth range and of the mth midrange are

$$L_w(t, m) = L_m(t) + {}_mL(-t); L_v(t, m) = L_m(t) + {}_mL(t).$$

If the initial distribution is symmetrical the semi invariant generating function of the mth range is twice the semi invariant generating function of the mth extreme value from above. The distribution of the mth range is skew, whereas the distribution of the mth midrange is of the generalized, symmetrical, logistic type. The even semi invariants of the mth midrange are equal to the even semi invariants of the mth range. For increasing indices m the distributions of the mth extremes, of the mth ranges and of the mth midranges converge toward normality.

6. **Statistics of Sensitivity Data, II.** Preliminary report. C. W. CHURCHMAN, Frankford Arsenal, and BENJAMIN EPSTEIN, Westinghouse Electric and Manufacturing Co.

In this paper a study is made of the distribution of the first two moments of sensitivity data as functions of the sample size. The chief results are briefly these:

(a) The distributions of $\bar{x}$ and $\sigma_{\bar{x}}^2$ (for definition of these functions, see "On the Statistics of Sensitivity Data," by the authors in the *Annals of Mathematical Statistics*, Vol. XV, No. 1) approach normality rapidly as functions of the sample size;

(b) $\bar{x}$ and $\sigma_{\bar{x}}^2$ are "almost" independent even for small sample sizes, thus justifying the use of Student's ratio in texts of significance for differences between two sample means.